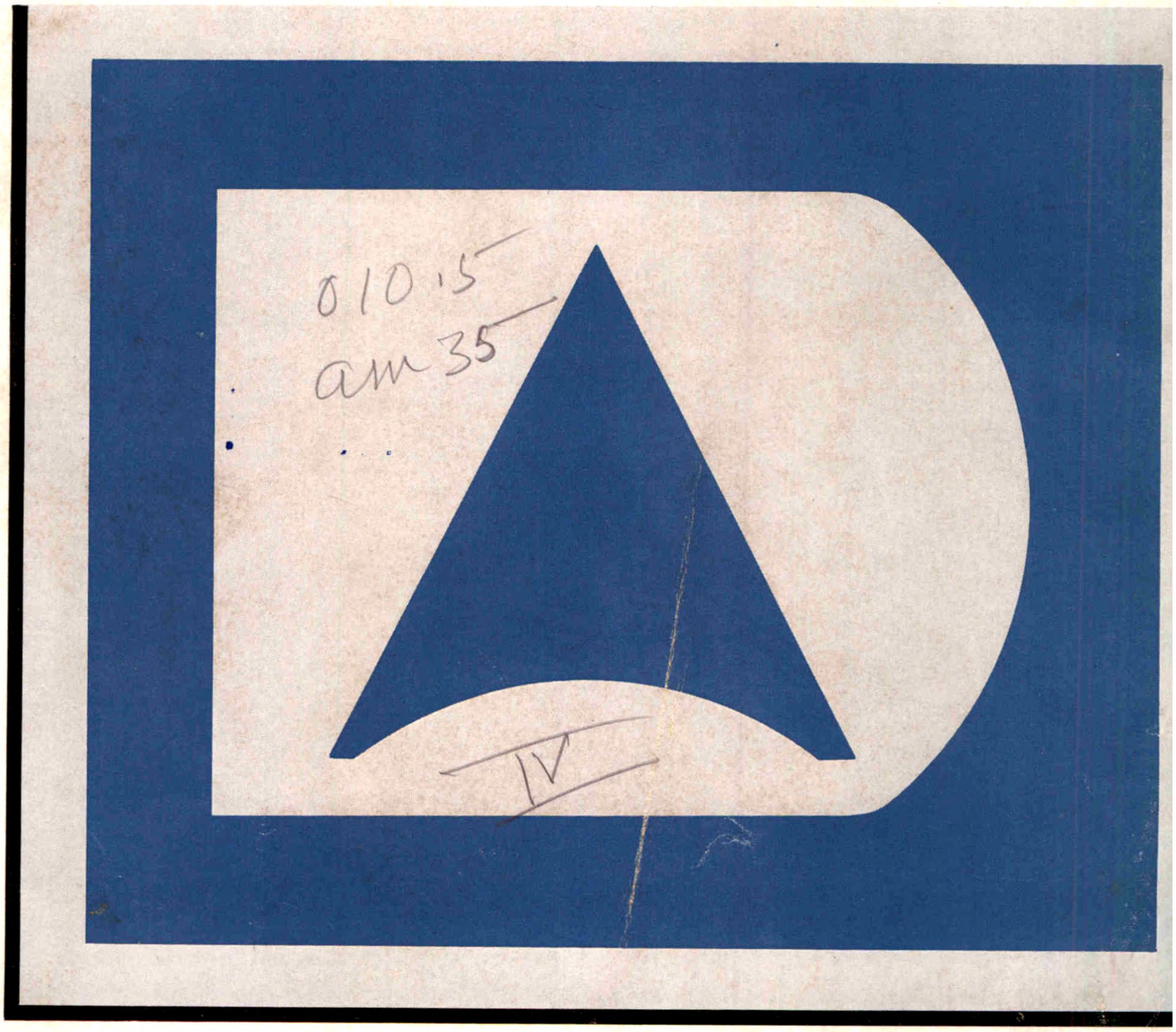
American Documentation

JANUARY
1966
Vol. 17, No. 1

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE



AMERICAN DOCUMENTATION

INSTRUCTIONS TO AUTHORS

American Documentation is a publication of the American Documentation Institute. It is a scholarly journal in the various fields in documentation and serves as a forum for discussion and experimentation. Papers already published or in press elsewhere are not acceptable. For each proposed contribution, one original and two copies (in English only) should be mailed to Mr. Arthur W. Elias, Editor, *American Documentation,* Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106. The manuscript should be mailed *flat* in a suitable-sized envelope. Graphic materials should be submitted with suitable cardboard backing.

TYPES OF MANUSCRIPTS: Three types of contributions are considered for publication: full-length articles, brief communications of 1,000 words or less, and letters to the editor. Letters and brief communications can generally be published sooner than full-length manuscripts. Books, monographs, and reports are accepted for critical review. Two copies should be addressed to the Review Editor, Dr. T. Hines, 54 North Drive, East Brunswick, New Jersey.

PROCESSING: Acknowledgment will be made of receipt of all manuscripts. *American Documentation* employs a reviewing procedure in which all mansucripts are sent to two referees for comment. When both referees have replied, copies of their comments are sent to authors with the Editor's decision as to acceptability. The refereeing procedure requires about 30 days. Authors receive galley proofs with a five-day allowance for corrections. Standard proofreading marks should be employed. Reprint order forms are forwarded with galleys.

FORMAT: All contributions should be typewritten on white bond paper on one side only, leaving about 1.25 inches (or 3 cm) of space around all margins of standard, letter-size (8.5 × 11 inch) paper. Double spacing must be used throughout, including the title page, tables, legends, and references. The first page of the manuscript should carry both the first and last names of all authors, the institutions or organizations with which the authors are affiliated, and notation as to which author should receive the galleys for proofreading. All succeeding pages should carry the last name of the first author in the upper right-hand corner (0.5 inch from the top) and the number of the page.

STYLE: In general, style should follow the forms given in the Style Manual for Biological Journals (SMBJ), published for the Conference of Biological Editors by the American Institute of Biological Sciences (1964).

TITLE: The title should be as brief, specific, and descriptive as possible. Vague and unrevealing titles may delay publication.

ABSTRACT: An informative abstract of 200 words or less must be included, typed with double spacing on a separate sheet. This abstract should present the scope of the work, methods, results, and conclusions.

ACKNOWLEDGMENTS: Financial support may be listed as a footnote to the title. Credit for materials and technical assistance or advice may be cited in a section headed "Acknowledgments," which should appear at the end of the text. General use of footnotes in the text should be avoided.

GRAPHIC MATERIALS: *American Documentation* requires finished artwork. Follow the style in current issues for layout and type faces in tables and figures. A table or figure should be constructed so as to be completely intelligible without further reference to the text. Lengthy tabulations of essentially similar data should be avoided.

Figures should be lettered in black India ink. Charts drawn in India ink should be so executed throughout, with no typewritten material included. Letters and numbers appearing in figures should be distinct and large enough so that no character will be less than 2 mm high after reduction. A line 0.4 mm wide reproduces satisfactorily when reduced by one-half. Graphs, charts, and photographs should be given consecutive figure numbers as they will appear in the text; however, figure numbers and legends should not appear as part of the figure, but should be typed double spaced on a separate sheet of paper. Each figure should be marked *lightly* on the back with the figure number, author's name, complete address, and shortened title of the paper.

For figures, the originals with two clearly legible reproductions (to be sent to referees) should accompany the manuscript. In the case of photographs, three glossy prints are required, preferably 8 × 10 inches.

ORGANIZATION: In general, papers should state the background and purpose of the study, followed by details of methods, materials, procedures, and equipment. Findings, discussion, and conclusions should appear in that order. Appendixes may be employed where appropriate for extensive lists, statistics, and other supporting data.

BIBLIOGRAPHY: Accuracy and adequacy of the references are the responsibility of the author. Therefore, literature cited should be checked carefully with the original publications. References to personal letters, abstracts of verbal reports, and other unedited material may be included. If an as-yet-unpublished paper would be helpful in the evaluation of a manuscript, it is advisable to make a copy of it available to the Editor. When a manuscript is one of a series of papers, the preceding member of the series should be included in literature cited.

CITATION FORMAT:

Order: Literature cited should be sequentially numbered as cited.

Authors: Give all authors with arrangement as follows:

Elias, A. W., B. H. Weil, and I. D. Welt

Titles: Give full titles of articles in English, indicating language of original as: (In Ger.)

Journals: Journal titles should be given in full.

MONOGRAPH AND SERIAL DATA: Should be presented in order as follows: Volume, issue number, pagination, and year. The issue number should be given in parentheses if journal pagination is not continuous from issue to issue. Pagination should be inclusive. Year of publication should be given in parentheses. An example is given below:

Bishop, D., A. L. Milner, and F. W. Roper, Publication Patterns of Scientific Serials, American Documentation, 16 (No. 2): 113–21 (1965).

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), three copies in the contributing membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts, Science Citation Index, Chemical Abstracts,* and *Documentation Abstracts.*

American Documentation is entered for second class mailing at Baltimore, Maryland.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 17, No. 1 JANUARY 1966

Published quarterly by AMERICAN DOCUMENTATION INSTITUTE and printed by Port City Press, Inc., 1323 Greenwood Road, Baltimore, Maryland 21208. Second class postage paid at Baltimore, Maryland.

Editorial

Documentation Abstracts

The Editor takes great pleasure in calling the attention of ADI members and *American Documentation* subscribers to the advertisement appearing on the back cover of the present issue. The publication of *Documentation Abstracts* as a separate publication, under the joint auspices of the American Documentation Institute and the Division of Chemical Literature of the American Chemical Society, had been a cherished dream of myself, Hans Peter Luhn, and many others for so long that it often seemed *only* a dream. It would be impossible to list all those who have lent support and encouragement to this effort, but the names of a few must be recited: Dr. Herman Skolnik, of the *Journal of Chemical Documentation*; Mr. Charles Bourne, long-time Editor of the Literature Notes Section of *American Documentation*; and Mr. John Markus and Miss Mary E. Stevens, of the ADI Publications Committee, have rendered yeoman service.

Documentation Abstracts, whose logo appears on the cover of this issue, will appear quarterly with issue Number 1 appearing in February. It is estimated that each issue will contain 500 abstracts from a unified coverage list representing the interests of ADI, the Chemical Literature Division, and the Special Libraries Association—Sci Tech Division. The latter group is represented in this project through the efforts of Mr. Charles Kip, who was Editor of "Documentation Digest." Mr. Kip has worked with the group designing the publication and has secured the continuing cooperation of the "Documentation Digest" abstracting group.

The Editor is happy to report that by action of the Council of the American Documentation Institute taken December 16, 1965, all members of the Institute will receive the new publication at no charge. Every effort will be made to continue this policy. However, members can help in this effort by promoting outside subscriptions to the new publication. The charge for a subscription in 1966 has been set at a minimal fee of $8.00.

Good luck to *Documentation Abstracts*!

A. W. Elias, Editor

An On-Line Technical Library Reference Retrieval System

In October 1964, Lockheed Missiles & Space Company (LMSC) started to experiment with an on-line reference retrieval system which uses a coordinate search strategy. Installation of the retrieval system was greatly facilitated by the existence of the LMSC on-line Automatic Data Acquisition (ADA) system which provided the vehicle for this application.

Experience with the current hardware-limited design has led to a more flexible "conversational" approach which is currently being implemented. The current system and the second-generation design using a "dialogue" are briefly described.

D. L. DREW, R. K. SUMMIT, R. I. TANAKA, and R. B. WHITELEY

Electronic Sciences Laboratory
Palo Alto Research Laboratories
Lockheed Missiles & Space Company
Palo Alto, California

Received 23 July 1965

• Introduction

Information retrieval is a vitally important problem and a very popular area for research. Consequently, there exists an information explosion concerning information retrieval resulting in many debates and discussions in information-processing circles. However, no general discussion will be presented in this report, since the system to be described here is designed to retrieve references, i.e., the names of documents, not to retrieve information. At present, the system does not concern itself directly with the major intellectual and practical problem in this area — that of resolving the conflicting requirements that a document be described briefly, and then that this brief description be adequate for later retrieval of that document.

Working reference retrieval systems involving the use of digital computers are no longer uncommon and have become highly sophisticated in their search strategies. Almost invariably, the procedure used is to accumulate a number of requests, code each in several alternative search formulations, and make a computer run with the batch. This method has the virtues of simplicity and economy, especially when the material to be searched can be loaded on a semirandom-access storage device (such as a disc or drum) and organized in an inverted file.

This method has several inherent disadvantages as well. The principal drawback is that of any batch-processed computer job: Once the job is submitted, the die is cast. Aside from some mechanical internal options, there can be little flexibility in the computer processing, and no monitoring of program execution. Turnaround time can also cause difficulties. When each successive refinement of a search request imposes a delay of a day or more, fewer refinements can be introduced. Another disadvantage is that in most batch-processing systems, an interpreter is required to translate the user's request into an acceptable system formulation so that the search can be made. In general, it seems that the more intermediaries interposed between the eventual user and the document collection, the greater the chance for distortion of the user's intention.

There is an alternative to this well-worked approach, and one which is just beginning to be explored. This is the possibility of retrieval via an on-line dialogue. This approach has been under study by Kessler of the Massachusetts Institute of Technology (1), implemented in a small experimental design by Salton of Harvard (2), and is realized in the current Lockheed working system.

There is a general feeling that on-line retrieval is the next major development, and represents the retrieval system of the foreseeable future (3 and 4). The basic reasons that the LMSC group endorses this view are (a) some sources of semantic distortion are eliminated by putting the user directly in contact with the system and its vocabulary; (b) the use of a dialogue permits the user to develop a highly complex search formula by a series of simple steps; (c) the result (negative or positive) of one search can be applied immediately to the next so that the overall search can be considered as a sequence of attempts. The intelligence and experience of the user of the system are involved not just once, in the framing of an inflexible question, but are continuously engaged until an acceptable result is achieved.

As far as can be determined, there exist, aside from those described here, just two other on-line retrieval systems. The SMART (2) system of Harvard University is unusual in a number of respects. It offers a large number of combinations of search strategies, a choice among a group of eclectic techniques. It has the refreshing quality of a design that uses many reasonable means to an end, rather than selecting one and discarding all competitors. The system embraces indexing as well as retrieval and provides a step-by-step definition of strategy. To date, it is an experimental tool and not an operating function of an existing library.

The other similar system, developed in conjunction with Project MAC at Massachusetts Institute of Technology (MIT), was reported by Kessler and collaborators (1) the same week the Lockheed system was first tested. These two independent developments resemble each other in many respects, but there are three major differences:

1. The MIT system is currently dealing with a specialized scientific collection (physics journals) as opposed to report literature in many fields.
2. The MIT development already uses the teletype terminals which represent the next step in the evolution of the Lockheed system.
3. The data base of bibliographic information in the MIT file was apparently designed for the experiment whereas the Lockheed design uses an existing machine-readable master catalogue card.

No description of methods of vocabulary control or translation facilities is given in the cited reference.

These differences have had great influences on search strategies. An example is the strong bias toward citation tracing in the MIT design, which seems much more appropriate to a journal than to a report collection. Other differences may be noted by using Reference 1 and the present report.

● Converse I: The Current System

The system which is now working at LMSC depends upon two other operating Lockheed systems. The first, MATICO (Machine-Aided Technical Information Center Operations), supplies the machine-readable data base. The second, ADA (Automatic Data Acquisition), supplies the on-line computer and its time-sharing monitor. The existence of these facilities permitted the design, development, and initial testing of the present system in a 6-week period. The accumulated product of the MATICO program will allow the inclusion in the retrieval system of the entire LMSC in-house report collection of 100,000 titles as soon as the system can accommodate them.

THE SYSTEM AS SEEN BY THE USER

The input device as a 1-card reader which has a series of levers permitting the input also of a 10-decimal-digit number. (As will be described later, this device is to be replaced in a later version by a more convenient mechanism.) The adjacent output device is a teletype printer. Beside the card reader is a file cabinet containing prepunched cards prepared for insertion in the reader. Each card contains a term or number describing one or more of the documents in the collection.

The types of information used as descriptors are: (a) personal author, (b) all significant title words, (c) corporate author, (d) all subject headings (an average of three per report), (e) contract number, (f) original report number, (g) secondary report number, and (h) date of publication. A file was also generated describing security classification, but was too broad for use with the present system.

Aside from these descriptor cards, two standard control cards are used, the NOT card and the END card. If descriptor cards are inserted with no intermediate control cards, they are considered to specify a logical-product search prescription. An intermediate NOT card negates the descriptor immediately following, and the END card informs the program that the search prescription is complete.

Instructions to the system user are:

1. Select a descriptor for the desired report from one of the categories of information, and find the corresponding inquiry card in the labeled card file drawers. If there is no such inquiry card, another descriptor should be selected which will relate to the desired information. Continue until your certain knowledge of the document or subject is exhausted. If it is desired to exclude some document references, each descriptor card using a term to be excluded from the document description is preceded by a NOT card. The inquiry card contains (left-to-right) the prefix categorizing the information on this card, a code number for the descriptor, the descriptor itself, and at the far right the number of documents in the file using this descriptor. (If this number is 1, clearly no further cards are needed to retrieve the unique document reference.)

2. Place the first inquiry card, with letters right-side up and facing the user, in the long card slot on the top of the machine. Enter the inquiry card by depressing the metal bar on the back of the card slot. Repeat the procedure for each new card input. When all the inquiry cards have been entered, insert an END card.

3. The response to the inquiry will be printed out on the teletype printer in one of three ways:

a. If 1–3 documents are found, the total information on an edited library catalogue card will be printed out for each document.

b. If 4–15 documents are found, only the corresponding document reference numbers will be printed out. In this case, there are two options for the next procedure: (a) "refine" the information request by inserting more descriptor cards, or (b) enter the document reference numbers by using levels 4 through 9 of the card reader. When an END card is then entered, the corresponding catalogue information will be printed out.

c. If more than 15 documents have been found, the printer will only write the number of documents that meet the request. In this case, the inquiry request should be refined.

4. The information request can be refined by adjusting the lever on the extreme left so that a "1" appears in the viewing window. Another card is then selected and entered. The subject area may be limited by inserting a NOT card before the area of information not desired, or further restricting descriptors may be entered until the total number of documents retrieved is less than 15. The teletype printer will respond after each card insertion, giving the information appropriate to the number of document references retrieved.

5. The information on a typical edited library catalogue card is as follows:

```
                        >
45931  GDC—U414—61—ØØ5  37P  UN>
GENERAL DYNAMICS CORP.  ELECTRIC BOAT DIV., GROTON, CONN.>
  DIGITAL  SIMULATION  OF  A  CONFORMAL  DIMUS  SONAR  SYSTEM.
    PHASE 1.>
ARNOLD, C.R.  FEB 61  C-SV>
DIMUS (DIGITAL MULTIBEAM STEERING)/SONAR SYSTEMS/HYDROPHONES>
AD—265 398//>
```

6. Four further responses may be made by the system. "NO REPORTS FIT YOUR INQUIRY" may be printed if there are no reports corresponding to a given combination of inquiry cards. "REPORT NO. 000000 IS NOT ON FILE" indicates that there is no report on file in the system with the requested reference number. This might be due to an error in keying-in the number on the variable levers. "ILLOGICAL 1ST TERM. PLEASE RESUBMIT." indicates that the request has started with an END or NOT card, neither of which is permitted as the first card. The phrase "END OF RESPONSE" follows each system response.

BEHIND THE SCENES

The retrieval files are made up from the information on library "master catalogue cards." This information is then entered on up to 44 key-punch cards per catalogue card. Two files are created and are linked by the accession number of the document. The first, the inverted file, is derived from an "explosion" of the catalogue master. In this operation descriptors corresponding to all of the categories mentioned previously are identified and separated, are labeled with the proper letter prefix, duplicates are combined, and the file is alphabetized within the descriptor category. This file is then output in a suitable form for editing with each card displaying a descriptor followed by the card or cards containing the corresponding document accession numbers. This card file is then edited manually. File entries whose descriptors are nonsignificant terms are purged (using a list of 585 such terms compiled by Bell Laboratories) and synonymous descriptors are combined by grouping them together.

The final stage of the inverted-file program reformats the edited information and outputs both an inquiry card deck and the inverted-file tape which will be loaded on the retrieval file. This step of the program assigns an arbitrary sequence number to each descriptor. If several descriptors have been identified as synonymous, each is assigned the same sequence number.

It will be evident that this design gives a tight control of vocabulary. Every term which can be used in the form of an inquiry card to query the system must exist within the inverted file, and vice versa.

The second file, the "catalogue file" of complete document descriptions, is a machine edited and reformatted version of the catalogue master. Editing eliminates redundancy, and reformatting prepares the information for output on the teletype printer.

OPERATING EXPERIENCE

The first realistic file used in the system was hand coded and consisted of 100 document references. Since then the file has been expanded twice: first to 1,600 and then to 8,000 references. An interesting sidelight is the growth in the number of distinct descriptors with respect to growth in file size. Some, such as report numbers, naturally increase more or less linearly. Some others are described in Table 1.

These numbers display some interesting tendencies. The corporate author file is almost complete at 8,000 references. In fact, at 3,000 documents, the file is nearly as large as at 8,000. The asymptote is roughly 1,000 corporate sources.

The number of distinct title words is also approaching saturation, but the collection is not yet large enough to yield a good estimate for the limiting value. The other three descriptor categories are, at the 8,000-document level, in a region of linear increase. It seems likely that the slope of the personal author line and the contract number line will continue with slight change as the collection increases. The subject headings, being combinations of terms, are potentially much more numerous than the individual terms; these will eventually level off, but the tendency is not visible as yet.

In summary it might be said that the available statistics indicate that three numbers, .26 contracts per

TABLE 1.

	File size		
	100	1,600	8,000
Authors	97	1,223	5,781
Title words	359	3,015	7,289
Subject headings	281	2,781	10,083
Contract numbers	53	522	2,200
Corporate sources	74	763	933

document, .71 personal authors per document, and about 1,000 corporate sources, can be used as estimates to characterize the entire collection. Although this describes a specialized information center, it is an interesting observation that the study of about 3,000 documents should provide these statistics for other collections.

USER REACTIONS

As the utility of the system is directly related to the number of documents represented, it was decided not to risk large-scale use (and possible user disappointment) until the most recent increase to 8,000 documents. The period since this increase has not been long enough to afford reliable statistics about the reactions of the users of the system.

However, some patterns have become evident. Users have grasped the rationale of the system quickly and easily and, in the cases observed, have used it effectively and been satisfied with the results. Two sources of complaint are the awkward card input device and the fact that the collection of documents is not larger. An interesting remark was made by several users to the effect that it made them impatient to wait 60 seconds while their catalogue cards were being teletyped, even though they felt they were saving hours or days of manual search. It would seem that an automated system, to be completely satisfactory, has to respond within a few seconds and should present output results at roughly a normal reading rate.

• Converse II: System in Development

The statistics displayed in the previous section, relating the growth in the number of descriptors to the number of documents represented, clearly indicate the existence of a practical upper limit to the size of the collection in the Converse I retrieval file. For 8,000 documents, 38,285 inquiry cards were generated.

One way to avoid the deluge of prepunched cards and simultaneously maintain vocabulary control would be to furnish a key-punch machine and an authority list at each inquiry station. This solution was rejected as awkward for the user, disturbing to other library patrons, and in other ways inferior to the system sketched below.

Converse II is designed around a teletypewriter inquiry station and the principle of allowing a "free" vocabulary at the start of the retrieval dialogue. The overall search divides itself into three phases. The first phase relates the user's vocabulary to the system vocabulary to find appropriate search terms, the second phase is a preliminary search to restrict the collection to a number small enough to analyze in some detail, and the third phase is a detailed search of the small file.

The first phase requires that different forms of the same term must be recognized. This demands a study of suffixes and the development of a practical program designed to identify the variants in most (say 90%) of the cases. Another aspect of the problem demands that a thesaurus be made available to the user, and that this thesaurus be related to a parallel wordlist: that of the content-descriptive terms derived from the collection. It is foreseen that each time a word is output from the thesaurus it will be compared to the wordlist, and if the program finds an identical or variant term as an actual descriptor, this fact will be indicated by the display of a number (the number of documents labeled by this term) after the output thesaurus term. The thesaurus word and the list variant are identified for the purpose of the search.

Another problem that arises in the first phase is that of partial identification. Provision will be made to relax the demand for exact matching when searching on personal and corporate authors. In the case of personal authors, the obvious change is to permit the use of the last name with one or no initials. In the case of corporate authors it will be necessary to incorporate a "thesaurus" to identify not only variant usages but to permit the use of initials, etc.

The second phase of the search is basically devoted to creating subfiles of the total file, and uses Boolean inverted-file manipulations to do so. The process might be compared to a "coarse sieving," with the purpose of eliminating the references which seem clearly irrelevant so that a detailed search of the remainder becomes feasible. One feature of such a process is that a subfile created by a given formula can be used as a unit in the next formula so that the terms used in its creation are combined and there is no need to repeat the process that created the file. This use of subfiles is clearly recursive. Another point is that these subfiles are not destroyed until the total search is at an end, so that partial backtracking is facilitated.

The third phase, that of serial search, will permit the use of dates such as "published after June 1962" and the use of security classifications as identifiers. Eventually this phase will be developed to a greater degree of sophistication; but there is a limit in the basic data set itself — almost all the information on the catalogue card is being exploited already. When abstracts become available to the system, many further refinements will be made. This is the area in which most future developments are to be expected, so an effort is being made to keep it flexible and open for new kinds of search tactics.

Plans have been made to place a terminal in the area where the library staff indexes incoming material. This terminal will provide access to the current thesaurus which is a locally modified version of the Engineers Joint Council retrieval thesaurus (5). the retrieval wordlist, and the corporate-source authority list, so that the indexing and retrieval vocabularies will become as similar as possible.

As a parenthetical note, it should be added that a further modification, Converse III, is under consideration. The major advantage of this future development will be the availability of a CRT output device to allow the retrieval program to become self-explanatory (a "tutorial" mode of operation) as well as to remove most of the restrictions on output imposed by the low speed of the teletypewriter.

References

1. Kessler, M. M., Ivie, E. L., and Mathews, W. D. 1964. The M.I.T. Technical Information Project — A Prototype System. *Proc. Am. Doc. Inst.*, 1.

2. Salton, G. 1964. A Document Retrieval System for Man-Machine Interaction. *1964 Proc. ACM.* Computation Laboratory of Harvard University.

3. Swanson, D. R. 1964. Design Requirements for a Future Library. *Libr. and Automation.* Library of Congress, Washington, D. C.

4. American Library Association. 1963. The Library and Information Networks of the Future. *USAF RADC Tech. Develop. Rep.*, No. 62–614.

5. *Thesaurus of Engineering Terms.* 1964. Engineers Joint Council, New York.

On Laws of Special Abilities and the Production of Scientific Literature

A quantitative estimate is made of the magnitude of the problem posed by the quantity of scientific and technical literature produced using improved estimation procedures, which produce more conservative estimates than those previously offered. The key is a method of estimation of literary productivity per man. The possibility that many contributions are presented in more than one publication is suggested.

LEROY H. MANTELL

Nasson College
Springvale, Maine

● Introduction

Recent work by de Solla Price (1) and Bourne (2) emphasizes the need to provide more definitive estimates of the literary productivity of scientists and engineers. It is possible that the so-called information explosion is not as serious as has been thought. In any event it is time that more powerful tools of analysis were brought to bear on the problem of estimating the volume of documentation because very real operational decisions depend on the accuracy of these estimates.

Previous writers have relied on estimates of the number of journals and a wide range of empirically arrived at averages of the number of articles per journal. Others have relied on estimates of the number of research reports associated with research and development contracts. The former method is handicapped because, as Bourne points out, we are not sure how many articles there are in a journal per year, nor which articles are eligible to be considered as technical literature. If we consider, as some have, the indexing process as sufficient evidence of contribution, we run in to the problem that there is duplication among indexes; that is, articles appear in more than one index.

In order to remedy the difficulty, some work done by Lotka and others in the field of laws governing special or unique abilities was reviewed. Lotka's law states that the number of people writing n papers in a lifetime is proportional to $1/n^2$. Unfortunately this does not help to forecast annually the output of technical papers, nor does it take account of highly prolific authors (3).

Fortunately there is another kind of mathematical generalization called the Poisson distribution which can be used, rather effectively, for our purpose. The Poisson may be used to estimate the probability of some event occurring during a period of time. That is, if L is the expected number of arrivals in a period of time, the probability of exactly n arrivals in the period is given by $L^n e^{-L}/n!$

To illustrate the application of this formula, Bortkiewicz's classic example concerning deaths from the kick of a horse in the Prussian Army is often selected, although needless to say many more up to date and equally as valid, but prosaic, examples are available.

Von Bortkiewicz collected information concerning the number of deaths which had occurred in a certain group of ten Prussian Army Corps over a 20-year period from 1875 to 1894 as a result of soldiers being kicked by a horse. He found 122 deaths recorded in the annual reports included in the study. The 200 reports in which the 122 deaths were recorded were distributed as shown in Table 1.

This kind of information precisely fits the Poisson esti-

TABLE 1.

No. of deaths per Army Corps per year	No. of reports
0	109
1	65
2	22
3	3
4	1
5	0

mating expression. And, as we will see, the number of learned papers or publications occurring in journals during a fixed time period also fits this expression. As a result, if we know the number of scientists and engineers, it is possible to estimate the number of papers that will be produced.

The results are based on tabulations of the number of learned papers and research results reported in a list of 29 abstracts and journals indexes. A value for L of .3 is found and a new procedure due to Cohen (4) is used to estimate this average value when the zero frequencies are unknown. It is then possible to estimate the probable output of research articles and reports of all United States scientists and engineers working in research and development. It will be seen that this method of forecasting research output gives results which are somewhat more conservative than others.

• Some Background

Measurement of literary productivity was approached by Lotka through the use of an index of all known articles to the year 1900 appearing in Auerbach's *Geschichtstafeln der Physik.* He also used the decennial index of *Chemical Abstracts* for the years 1907–16, counting the number of names against which there appeared 1, 2, 3, etc., entries for those names beginning with A and B (a random sample) (5).

Dresden in 1922 tabulated the number of contributions of learned papers to the 25th anniversary meeting of the American Mathematical Society (6). Neither Lotka nor Dresden concern themselves with the larger universe of noncontributors and the latter is forced to group his information in a specific way in order to obtain a suitable mathematical representation. Certainly if a formula is to be used which will estimate the probability of a person producing a learned paper, it should not be made to depend for its accuracy on the way in which information is classified. Also, as has been noted in the case of Lotka, but not shown here, the form of mathematical representation should not change merely because the time period covered has changed.

We are, however, indebted to both Lotka and Dresden, and using their technique the numbers of contributions per author listed in the Office of Naval Research Human Engineering Bibliography 1959–1960 were tabulated with the results shown in Table 2 (7).

TABLE 2. Contributions per Author-ONR Human Engineering Bibliography

No. of titles listed	No. of authors mentioned
1	1,718
2	332
3	109
4	48
5	31
6	10
7	3
8	2
13	1
16	1
	2.255

In other words, 1,718 out of 2,255 authors listed produced one article or report only, 332 authors produced two each, and so on.

When Lotka and Dresden attempted to fit mathematical expressions to data of this kind, they neglected to account for the number of persons who produced no output. For our purposes, measurement of this class is of vital importance, because many persons work diligently without producing anything of note to be shown within a given period of time. This factor becomes of even more importance the shorter the time period to be measured.

• The Estimating Procedure

The question now to be answered is, how many produced no reports? With such an estimate, the number who produced reports plus those who did not, gives us the total number in the universe, and deriving such a parameter we can estimate the number of papers a given population of potential contributors is likely to produce.

Fortunately Cohen (8) has provided an answer. Using his tables for estimating the mean of a Poisson distribution with the zero value absent, we are able to restate the distribution as shown in Table 3.

We see that the use of the procedure has resulted in an estimate of over 2,000 noncontributors. Or, reversing our procedure we can say that had we started with a population of 4,353 specialists in Human Engineering, we could have estimated 2,255 papers as the total output of this group.

Obviously the procedure is only as valid as the method

TABLE 3. ONR Human Engineering Bibliography-Contributions per Author — Zero Frequency Supplied

No. of titles	Authors	%
0	2,098	48.2
1	1,718	39.5
2	332	7.6
3	109	2.5
4	48	1.1
5	31	.7
6	10	.2
7	3	—*
8	2	—*
13	1	—*
16	1	—*
	4,353	100.0

* Less than .1%.

of computing the zero frequency. Equally also, such a procedure must have some error limits associated with it; i.e., a range of values within which there is a strong probability a correct estimate lies.

The important parameters here are the averages because essentially what the process does is to amend the average of the distribution calculated without the zero frequency in such a way that a new average is calculated as though there were a zero frequency. In the instance of the ONR data, the average output per man is 1.41 articles. The calculated L or average of the Poisson distribution which assumes there is a zero class, and which matches this mean of 1.41, is .7474. It develops also that the three standard deviation limits for this estimate are .7543 and .7405 which means that although the estimate of authors given in Table 3 is 4,353, an error of more than .4% or 16 authors either way is extremely unlikely. That is the error of the estimate is very small.

• The Time Effect

A similar computation, performed on data contained in Communications of the Association for Computing Machinery Author Index 1958–1961, produced Table 4.

The significance of the above table lies in the fact that the proportion of the frequencies in the zero class is less than the proportion in the class having one title each. In other words, the increased length of time, in this case 4 years, is apparently responsible for increases in literary productivity. The Human Engineering Bibliography covered one year's work; the ACM index covered four year's work. The proportion of authors having no contribution in the first case was about 48%; in the second 31%.

With this as a clue, a number of tabulations of author indexes of learned periodicals and journals were made. The tabulations covered all time periods, from biweekly indexes of active research aids such as *Chemical Abstracts* to the 50-year record of contributions to the *Quarterly Journal of Economics*. The broad outline of system behavior was first established, then an intensive study of annual productivity was undertaken.

In all, 3 biweekly, 2 semimonthly, 3 monthly, 4 bimonthly, 1 semiannual, and 8 annual references, plus 5 references covering periods longer than annual, were used to establish the broad outline of the behavior of productivity over time. The intensive study of annual productivity covered 29 references. A listing of references is given in Appendix A to this article.

Table 4. ACM Author Index: Contributions per Author 1958–1961 — Zero Frequency Supplied

No. of titles	No. of authors	%
0	44	31.2
1	54	38.3
2	27	19.1
3	13	9.2
4	2	1.4
9	1	.8
	141	100.0

The technique amounted to counting the number of mentions opposite an author's name over a sample number of pages of the index. Joint authorship of an article was counted as single authorship and subsequent mentions under joint authors were ignored. Since actually each author should have received credit for less than a complete report, the final results will tend to be overstated because the percentage of contributors producing one report will be larger than it should be. To some extent, however, this may be compensated for by the fact that joint authorship produces more reports. That is, a team of four scientists may produce six reports under joint authorship. This would be counted as one author producing six reports and the upper ranges of the distribution would then tend to be overstated.

The sample selection was strictly random, although there was a preference for the use of the alphabet as a basis for selection. In other samples a sequence of pages was selected at random.

Table 5 contains the result of taking samples of varying sizes from different abstracts and subjecting them to the Cohen technique. Where the number differs from that referred to earlier it is because samples from different time periods were combined for identical references.

The trend follows the intuitive belief established earlier, namely, the longer the period of time covered by the index, the greater the probability of a contribution to research, and the shorter the time period, the greater the likelihood of no contribution during the time period.

• Making the Most of a Contribution

Since one of the aims of this research was to produce an annual estimate of literary productivity of scientists

Table 5. Relationship of Per Cent Noncontributors to Time Period Covered by Abstract

Time period	Noncontributing %
Biweekly (2)	90.9, 79.9
Semimonthly (1)	76.0
Monthly (2)	89.4, 80.6
Bimonthly (3)	82.0, 77.4, 75.3
Semiannual (1)	56.1
Annual (8)	56.2, 54.5, 49.5, 48.2 43.6, 40.6, 38.8, 36.3
Two years (1)	48.2
Four years (1)	31.2
Five years (1)	27.4
Fifteen years (1)	36.0
Fifty years (1)	8.4

TABLE 6. Per Cent in Zero Frequency Class, Annual Journal Indexes only

Zero frequencies %	No. of journals
91–100	1
81– 90	6
71– 80	9
61– 70	2
51– 60	1
41– 50	1

and engineers, journal literature indexes for one-year periods were studied intensively. Abstracts were not used at this time because of the possibility of duplication of articles. This simple precaution resulted in a rather interesting corollary to the main conclusion of the work, as we shall see shortly.

Twenty journal indexes were examined and sampled, and frequency distributions obtained as before. The data were then subjected to the Cohen procedure with the results shown in Table 6.

By comparison, the data in Table 5 for abstracts only, publishing on an annual basis, gives a range of 56.2% to 38.8%, with an average for the zero class of 50%. But the average of the above frequency distribution is 75.7%. The difference can be seen quite clearly in Table 7 where specific fields are given.

In the range covered by the five listed above, about 76% of the authors will submit no papers for journal publication during a year. Those that do, are likely to prepare on the average of two articles on any annual contribution.

These conclusions are arrived at by the following means. The average proportion of scientists contributing to journals is 24%. The average proportion contributing to abstracts is 50%. But abstracts cover all journals in their field. Then two mentions in the abstract, with only one mention in a particular journal would mean that a scientist making a research contribution is likely to prepare articles in more than one field, or more than one journal.

When tabulating from each journal, a scientist making a contribution is counted once. If a scientist has made contributions to two journals, he is counted as having made one contribution for each medium. In the abstracts, however, he is counted twice; one for each article.

TABLE 7. Per Cent in Zero Class for Specific Fields: Journal Indexes and Abstracts

Subject	Per cent in Zero Class	
	Journal index	Abstract
Ceramics	89.7	54.5
Chemistry	69.5	38.8
Psychology	85.7	56.2
Physics	62.6	40.1
Aeronautical Eng.	73.6	49.5

The further conclusion we draw is that we must look to the journal indexes rather than the abstracts for our productivity factor. Instead of the approximate 50% zero productivity factor obtained from abstracts data, we must look at the 75.7% factor obtained from the review of the journals, if we wish to get a true measure of productivity.

• Preparing the Final Estimate

The actual data obtained from a review and tabulation of the annual author indexes of learned or technical journals amounts to samples drawn from the universe of all such journals in all time periods. The details are shown in Table 8. The frequency totals obtained are given in Table 9 with the zero class supplied by computation.

The expected proportions are those which would have been obtained had the Poisson distribution been used instead of the actual data. The agreement is very close with chi square computed at 1.62. Such a value could have arisen more than 99 times out of 100 due to chance alone. The actual and expected results are shown also in Fig. 1.

• Some Conclusions

If we apply the productivity schedule listed above to the National Science Foundation estimates of the numbers of scientists and engineers in research and development in the United States, it should be possible to provide an estimate of the amount of significant technical literature being produced annually.

In April 1962, the NSF (National Science Foundation) reported the following numbers of scientists and engineers in research and development in the United States (9):

1954	223,200
1958	327,100
1960	387,000

By using several estimating methods which are essentially extrapolations from past growth rates, the following forecast of the number of scientists and engineers in research and development (excluding supervisory positions) can be made (10): [1]

January	1962	431,000
Average	1963	435,000
"	1965	480,000
"	1970	649,000

A straight line fitted approximately to the three estimates for 1954, 1958, and 1960 is given in Fig. 2. The

[1] Based on National Science Foundation data.

TABLE 8. Samples from Journal Author Indexes

Journal	Frequency of contribution per year						
	0[a]	1	2	3	4	5 & over	Total
Psychol. Bull.	337.0	30.5[b]	.7[b]				368.2
J. Am. Ceramic Soc.	1114	120	7				1241
J. Consult. Psychol.	772.0	104.7[b]	6.3[b]	.7[b]			883.7
J. Sci. Instrum.	734	103	8				845
Nucleonics	1193	191	15	1			1400
Psychometrica	150.0	28.5[b]	2.0[b]	.5			181
J. Appl. Psychol.	310	66	6	1			383
Trans. ASME	1099	252	23	5			1379
J. Phys. Chem.	467	106	12	1			586
Analytical Chem.	406	96	9	2			513
Bull. Am. Math. Soc.	338	87	10	1			436
J. Aerospace Sci.	362	114	11	5			492
J. Appl. Phys.	364	117	18	1	1		501
J. Chem. Phys.	252	87	11	2	1		353
J. Opt. Soc. Am.	290	101	16	3			410
Trans. Am. Geophys. Un.	374	137	14	1	3	1	530
J. Acoust. Soc. Am.	234	101	14	5	1		355
Phys. Rev.	333	172	27	7	5		544
Am. J. Math.	78	43	7	2	—	1	131
J. Am. Chem. Soc.	125	112	12	6	1	5	261
Total	9332	2169	229	44	12	7	11793
Per cent	79.1	18.4	1.9	.4	.1	—	100.00

[a] Calculated.
[b] Represents average of two different year's indexes.

line passes through 420,000 at January 1962, 470,000 for June of calendar year 1963, and could well produce an estimate of 640,000 for 1970. It would appear that most estimates of the number of scientists and engineers are based on linear growth rates.

Now clearly the number of scientists and engineers assumed must be in place at least one year in some assignment or another in order to be productive in their line of work. We therefore assume that the number of personnel in place in July 1963 would be giving rise to publications in July 1964 — certainly not prior to that date, and most probably later than that date.

There are certain advantages also in working with estimates for the year 1963. First, the forecast of the number of scientists and engineers has less opportunity to suffer effects of errors in extrapolation, and second, we may possibly be in position to verify the accuracy of the final result. It is necessary to remember, of course, that the publications estimate will be for 1964.

TABLE 9. Contributions per Author: Samples of Annual Journal Author Indexes — Zero Frequency Supplied

No. of titles listed	No. of authors mentioned	% Actual	Expected distribution
0	9,332	79.1	74.1
1	2,169	18.4	22.2
2	229	1.9	3.3
3	44	.4	.3
4	12	.1	.1
5	7	—*	—*
	11,793	100.0	100.0

* Less than .1%.

Working with an estimate of 465,000 potential authors as of July 1963, the probable annual productivity is calculated in Table 10.

Compared with the rather large numbers that have been offered by reputable and conscientious researchers in the field, it would appear almost obligatory to apologize for the small size of the estimate.

On the other hand, it does represent an average of 3.32 scientists and engineers per research paper, taking into account that 74.1% of all of them will produce no papers during the year. By comparison the average number of basic research scientists and engineers per basic research paper in 1959 is reported by the National

TABLE 10. Calculation of Literary Output of Scientists and Engineers in Research and Development during 1964

Expected proportion	No. of authors	Contributions per author	Total articles
22.2%	465,000	1	103,230
3.3	465,000	2	30,690
.3	465,000	3	4,185
.1	465,000	4	1,860
Total expected number of articles			139,965

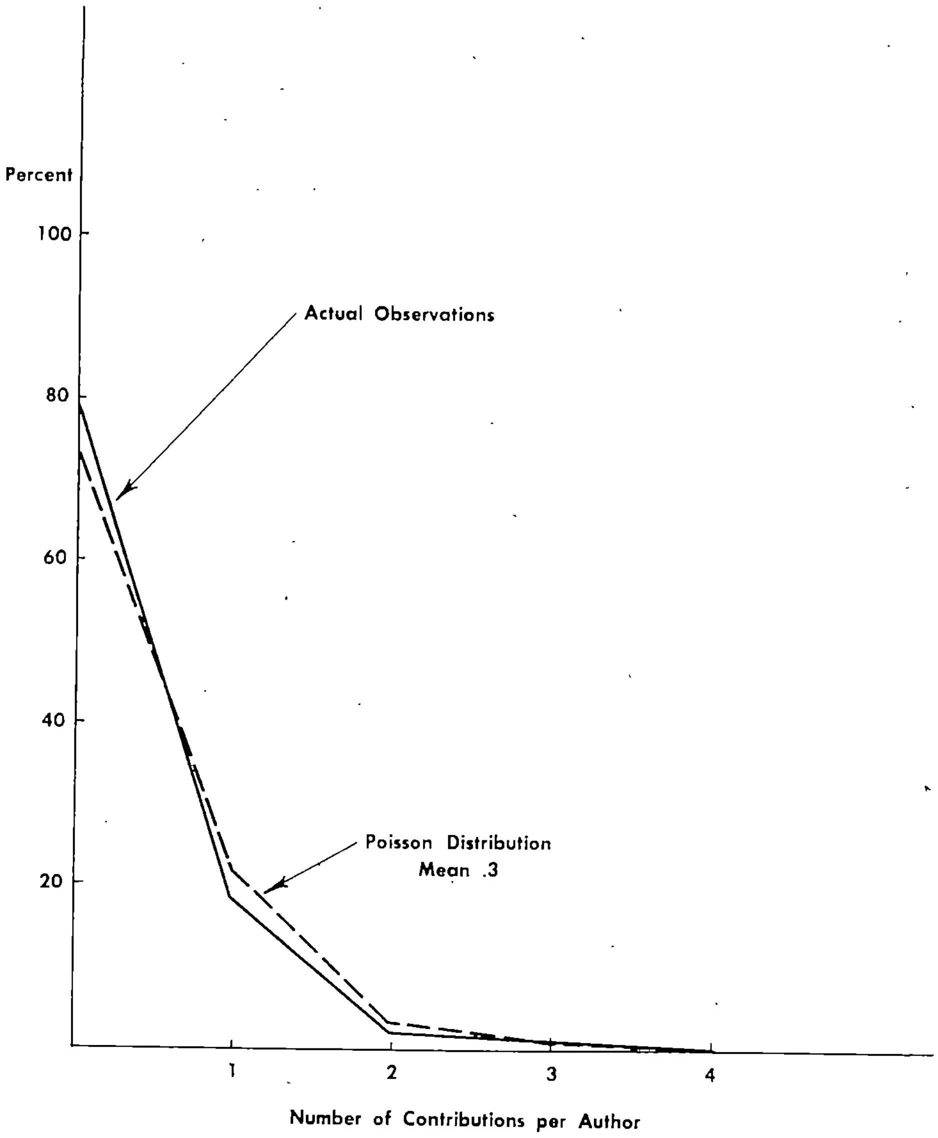


FIG. 1. Per cent contributing and number of contributions per author annually — journal indexes.

Science Foundation as 2.4, a not unexpected difference (11).

• Comparison with World-Wide Productivity Estimates

If we consider the productivity of scientists and engineers using the results obtained from sampling abstracts rather than the more conservative journal estimates, our productivity factor now becomes 1.46 scientists and engineers per paper written, a little more than twice the former estimate obtained.[2]

If this average productivity is valid, the 480,000 scientists and engineers will produce 328,000 articles and reports in the United States in 1965.

By comparison, Kent (12) forecasts a total world output of technical literature in the same year of over 900,000.

When we compare these estimates, we see that the United States will produce over one third of the world's technical literature in 1965. Of this number the Department of Defense has financed possibly 174,000 and of these 75,000 will represent actual research results, rather than reworkings or reinterpretations.

• Summing Up

Briefly then, what has been shown is that despite frequent references in the press and elsewhere concerning

[2] Based on the following schedule of productivity:

0	40.5%	4	1.3%
1	38.2	5	.5
2	8.2	6	.1
3	2.1		

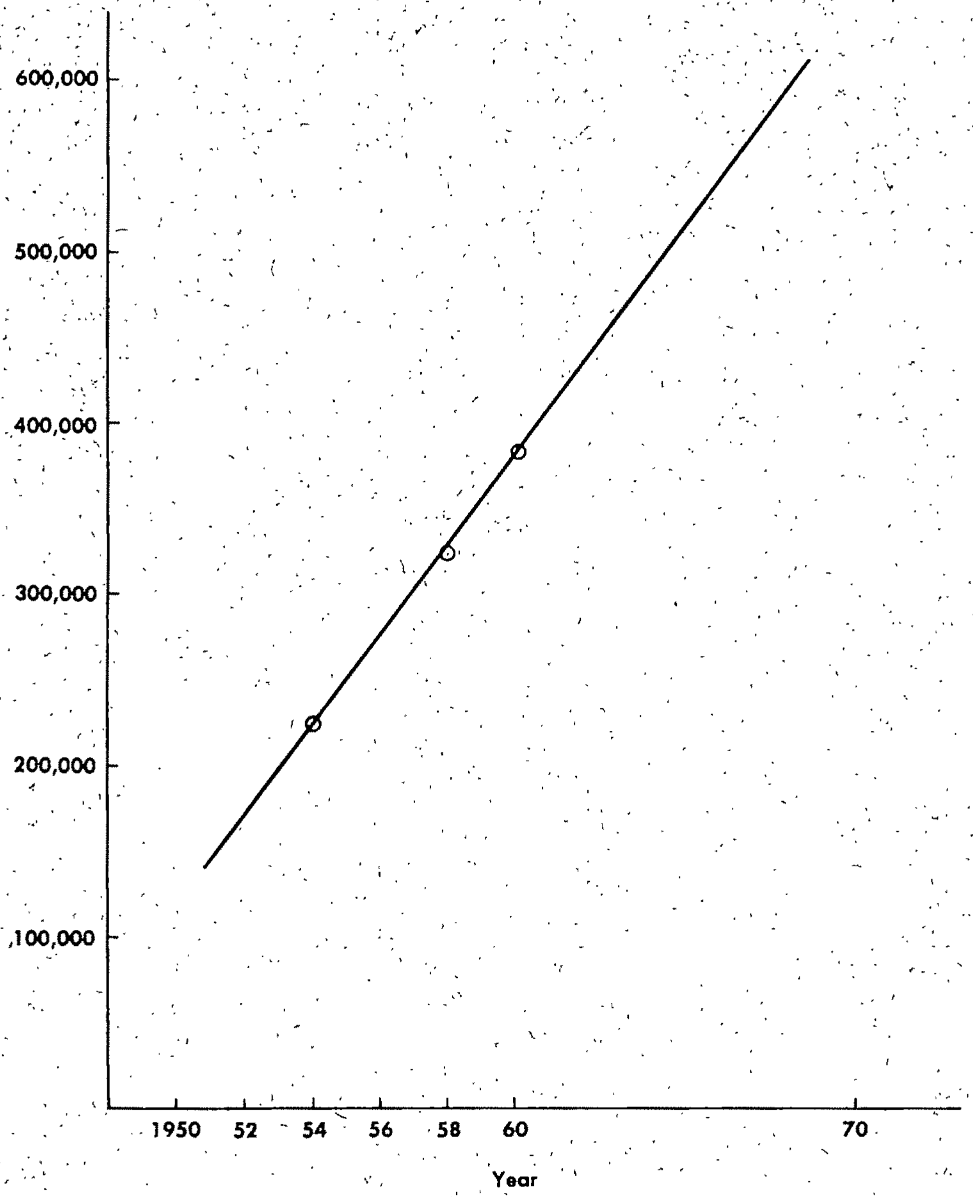


FIG. 2. Number of scientists and engineers in the United States in research and development.

an exponential growth in technical literature, there does not seem to be any evidence to this effect. The growth of the numbers of scientists and engineers seems to be linear, not exponential, according to reliable estimates. Secondly, analysis of actual productivity applied to these estimates, fails to indicate anything like the volume of information other writers have predicted would occur.

Of course the procedure does assume a constant average productivity per investigator based on journal output during the period 1960–61. Should this average productivity be found in later years to have risen, then estimates of annual output would also have to increase. Certainly a more exhaustive investigation of the whole field is indicated by these preliminary results.

References

1. DE SOLLA PRICE, D. J. 1963. A Calculus of Science. *Internat. Sci. Technol.*, March: 37–42.
2. BOURNE, C. P. 1962. The World's Technical Journal Literature: An Estimate of Volume, Origin, Language, Field, Indexing, and Abstracting. *Am. Doc.*, April: 159–168.
3. BOURNE, C. P. *Op. cit.*
4. COHEN, A. C., JR. 1958. Estimating the Poisson Parameter from Truncated Samples with Missing Zero Observations. Technical Report No. 15, University of Georgia, Department of Mathematics, No. DA-01-009-ORD-463, Department of the Army.
5. DAVIS, H. T. 1941. *The Theory of Econometrics*. Bloomington: Principia Press, pp. 49–50.

6. DRESDEN, A. 1922. A Report on the Scientific Work of the Chicago Section, 1897–1922. *Bull. Am. Math. Soc.*, **28:** 303–307. (See also Davis, *op. cit.*, pp. 48–49.)

7. TUFTS UNIVERSITY. 1961. Institute of Psychological Research, No. ACR-69, Office of Naval Research. Report, October.

8. COHEN, A. C. *Op. cit.*

9. NATIONAL SCIENCE FOUNDATION. 1962. Review of Data on Research and Development. Report No. 33, NSF 62-9, p. 6.

10. NATIONAL SCIENCE FOUNDATION. 1963. Profiles of Manpower in Science and Technology. Report, NSF 62-23.

11. NATIONAL SCIENCE FOUNDATION. 1961. Publication of Basic Research Findings in Industry 1957–1959. Report, NSF 61–62, pp. 28–29.

12. KENT, A. 1962. Resolution of the Literature Crisis in the Decade 1961–1970. *Res. Management*, January.

APPENDIX A

SOURCE OF DATA

Biweekly

Technical Publications Announcements, 27 September 1962, **2**(13), NASA, A–L inclusive, M–Z inclusive.

Chem. Abstr., 6 August 1962, **57**(3), American Chemical Society, A and B.

Semimonthly

Nucl. Sci. Abstr., Personal Author Index; 31 August 1962; U. S. Atomic Energy Commission; A–C inclusive, Q–Z inclusive.

Monthly

Index Med.; April 1962; **3**(4): National Library of Medicine, U. S. Department of Health, Education, and Welfare; pp. A415–A425 inclusive, pp. A494–A502 inclusive.

Solid State Abstr., 1962, **3**(2), Cambridge Communications Corp., Abstracts No. 14347–14694.

Bimonthly

Astronaut. Inform. Abstr., Reports and Open Literature, August 1962, **6**(2), Abstracts 60, 308–60, 603. Jet Propulsion Laboratory, California Institute of Technology, Pasadena. Author Index Nos. 60,001–60,603.

Psychol. Abstr., April 1962, **36**(2), American Psychological Association, Inc., Washington, D. C. Author Index A–D inclusive.

Psychol. Abstr., June 1962, **36**(3). Author Index, A–D inclusive.

Biol. Abstr., January–February 1961, **36,** University of Pennsylvania. Author Index, A–B inclusive.

Semiannually

Nucl. Sci. Abstr., Semiannual Index, 30 June 1962. January–June 1962, **16**(12B), U. S. Atomic Energy Commission. I, J, and part of K.

Annual

Phys. Abstr., Science Abstracts Section A, **63,** Author Index Number, 1960, pp. 2101a–2108b, 2277a–2281a.

———, 1948, **52,** A only.

Aeronaut. Eng. Index, 1957. Institute of the Aeronautical Sciences, New York, 1959. 547 items from beginning of A. 347 (all) P, and 546 from end.

Psychol. Abstr., Annual Index Number, December 1952; **26**(12), A–Bion inclusive and L–Mayer-Gross.

Chem. Abstr., Author Index, 1961, A–page 21, B–beginning page 73, pp. 253–261 inclusive.

Analyt. Chem., 1961, **33,** American Chemical Society, Washington. Author Index, pp. 1971–1980, A–D inclusive.

Trans. ASME, January 1958, **80,** containing Index to ASME transactions, 1957, **79,** pp. SR 133–SR 145, A–S inclusive.

———, containing Index to *Mechanical Engineering,* January–December 1957, **79,** pp. SR 105–SR 131.

J. Am. Chem. Soc., 1961, **83.** Author Index, pp. 5053–5108, A and B.

J. Am. Ceramic Soc., 1961, **44,** Columbus. Author Index, 1 December 1961, **44**(12).

Ceramic Abstr., 1 December 1961, American Ceramic Society.

J. Acoust. Soc. Am., December 1960, **32**(12). Author Index, **32,** p. 1722, A–D inclusive.

Bull. Am. Math. Soc., January–December 1961, **67.**

J. Chem. Phys., December 1961, **35**(6), pp. 2274–2285. Author Index, **35,** A–D inclusive.

J. Phys. Chem., 1960, **64.** Author Index, A–D inclusive.

J. Scient. Instrum., 1961, **38.** Index, A–L inclusive.

J. Appl. Psychol., 1961, **45.** Author Index, all.

J. Appl. Phys., December 1961, **32**(12). Author Index to Vol. 32, A–C inclusive.

J. Opt. Soc. Am., December 1960, **50**(12). Author Index, **50.** All.

J. Aerospace Sci., December 1961, **28**(12). Index, **28,** pp. 1000–1007. Author Index, A–G inclusive.

Phys. Rev., 15 December 1960, **120**(6). Author Index, **117–120,** covering the year 1960, A–C inclusive.

Psychol. Bull., 1960, **57.** Table of Contents.

———, 1961, **58.** Table of Contents.

Psychometrica, 1959, **23.** Index.

———, 1959, **24.** Index.

J. Consult. Psychol., 1959, **23.**

———, 1960, **24.**

———, 1961, **25.** Table of Contents.

Nucleonics, December 1960, **18**(12), Author Index. January–December 1960, **18,** pp. 158–160.

Transactions — American Geophysical Union, 1960, **41.** National Academy of Sciences, National Research Council, Washington, D. C.

Am. J. Math., 1951, **73,** Johns Hopkins Press, Baltimore.

Human Engineering Bibliography 1959–60. Report, October 1957, Tufts University, Institute for Psychological Research, Human Engineering Information and Analysis Service, Report No. ACR-69, Office of Naval Research.

Five Years

Psychopharmaca, a bibliography of Psychopharmacology, 1952–57. National Library of Medicine, Public Health Service, U. S. Department of Health, Education, and Welfare.

Public Health Service Publ. No. 581, Public Health Bibliography Series No. 19. Anne D. Caldwell, M. D., Ed. U. S. Government Printing Office, Washington, 1958. A and B.

Fifteen Years

Bibliography on Shock and Shock Excited Vibrations, January 1958. I. N. Brermnn, Ed. Engineering Research Bulletin No. 69, College of Engineering and Architecture, Pennsylvania State University. Covers 1938–56. A–E inclusive.

Fifty Years

Quart. J. Econ., Index 1886–1936. Harvard University Press, 1936. A–F inclusive.

Indexing Problems and Some of Their Solutions

This paper concerns problems of redundancy and inaccuracy in the indexing of technical information, illustrated by the coordinate indexing scheme employed in the NASA Information System. A proposal is made for the elimination of "panacea" or "catch-all" terms, and a rule for uniform grammatical negation is given. The effects of synonyms, antonyms, and negations on the overall efficiency of the information system are illustrated. Merits are discussed and rules are given for indexing under acronyms whenever possible. Finally, the concept of a "pictorial thesaurus" is proposed to exhibit hierarchy and connectivity of terms as an aid to indexing and retrieving of information.

LOUKAS LOUKOPOULOS

Center for Application of Sciences and Technology
Wayne State University
Detroit, Michigan 48202

• Introduction

The use of the computer, although a boon to rapid location of information, has stunted the growth of indexing techniques which promote optimum effectiveness of retrieval in a viable information system. The ease of information location contributes to redundant indexing by encouraging the assignment of many descriptors to a document. On the one hand, this facilitates the retrieval operation by minimizing the necessity of a rigorous presearch hunt for the exact terms which characterize the search topic. On the other hand, it often creates the need for reanalysis after retrieval because of the large number of irrelevant items printed out. Moreover, the cost of the time spent in screening the yield of documents for relevancy is an important factor, since in general, the searcher's or user's time is much too valuable to be spent in this way.

The principal requisites for an information-retrieval system are predictability and relevancy of the retrieved information. Unpredictability fosters doubt as to the completeness of an information-retrieval task. A consequence of this is more searches and an increase in output cost. If the literature searcher, in an effort toward completeness, engages in prolific machine searching within a given problem area, he will undoubtedly be confronted with redundancy in the retrieved information. The strategy underlying the retrieval of information from a computer-centered system can be very difficult to formulate. This is because of the multiple connotations of many descriptors, and the inevitable human inconsistencies in distinguishing among a given set of eligible terms under which a document may be indexed. The result is often an information system that is neither sufficiently accurate, nor sufficiently predictable. By careful procedures the experienced searchers can avoid many of the pitfalls of inaccurate or redundant indexing, but not without spending extra and valuable time.

This paper will consider only questions and problems on the indexing aspect of an information system, since indexing is the first step to accurate and useful information retrieval. To some extent this approach to indexing will be made from the standpoint of the grammatical structure and logical implications of descriptive terms. One reason for taking this approach is to obtain certain simple rules whereby the descriptor's grammatical structure can be used to differentiate among varieties that can be described by the term.

For purposes of illustration, this paper will draw upon the system model of the National Aeronautics and Space Administration (NASA) which services Wayne State University's Center for Application of Sciences and Technology (CAST). It will concentrate on the machine term vocabulary aspect, with occasional digressions into searching techniques. The attempt is to consider only a few aspects of the indexing operation, rather than to exhaust the whole topic. Hopefully, the results of this analysis will help in the design of simpler and more concise user-oriented information systems.

• The Information System

An information system is comprised of a computer, a storage element, a machine term vocabulary (MTV), a searcher, and a user. The MTV is a collection of all terms under which the information mass has been indexed and stored. The searcher is the person who formulates the user's question, in a language governed by MTV, and presents it to the computer. The user is the person who generates the question and who will ultimately use the retrieved information to accomplish a specific task.

Only the basic concepts of set theory are relevant to this discussion. Of particular use are the concepts of set, subset, the union (+), intersection (×), and complementation (c) of sets. The set-theoretic approach is taken for two reasons. The first one is based on the computer's ability to be programed to satisfy Boolean relations. The second reason is that each MTV term can be considered as a set whose elements comprise the documents indexed under the term.

In general, MTV terms can be divided into two mutually exclusive types. The simplest type of term is a single word such as AIR, HEAT, SOUND, WORK, etc. The other type of term is comprised of a series of words such as BUBBLE CHAMBER, CROSSED FIELD AMPLIFIER, PULSE WIDTH MODULATION, PLASMA ARC METAL SPRAYING, etc. For the present it will suffice to note that the second type of term is of the form AN, AAN, AAAN . . . where A and N stand for Adjective and Noun, respectively.

The MTV is a necessary link between searcher and computer because only those terms found in MTV are meaningful to the computer. In addition to placing a limit on the number of computer-understood terms, MTV actually defines the information mass as a function of the totality of its entries. This defining of the information mass is rather narrow because the index terms reflect the terminology of the author and to a lesser degree the terminology of the abstracter and indexer of each document. Oftentimes these terminologies do not coincide with the user's and effective communication among user, searcher, and computer, breaks down. For this reason, many reference works such as technical dictionaries and thesauri are essential to every presearch analysis. Some of the most useful references are *Thesaurus of Engineering Terms* (*EJC Thesaurus*), *Euratom Thesaurus, Space Age Dictionary, Van Nostrand's Scientific Encyclopedia,* and *Webster's New International Dictionary.*

In all that follows, it shall be assumed that an abstract of a document has been given for indexing. It will be further assumed that the abstract has been furnished by the author of the document, or generated by someone who is knowledgeable in the field about which the document relates.

• Panacea Terms

The foremost problem associated with the information system defined in this paper is the existence and rate of growth of "panacea" or "catch-all" terms. With particular reference to the NASA system, a *panacea* term is defined to be that MTV entry under which at least 1,500 abstracts have been indexed. More generally, *panacea* terms are defined to be those, which, through either their frequency of usage or broadness of scope, possess little or no definitive power. Examples of such terms are: AIR, ATMOSPHERE, DATA, HEAT, WORK, etc.

Panacea terms are detrimental to the effective operation of the information system for many reasons, the most important being unpredictability and irrelevancy of search results associated with their use. For example, under present indexing schemes, the term WORK has abstracts indexed under it that deal with such diverse topics as "working sleigh dogs," "work hardening of metals," "thermionic work functions," and many more. A searcher wanting only information dealing with "thermionic work functions" would have two search alternatives. The first would be to search the computer via the intersection THERMIONIC × WORK × FUNCTION/S. The second would be to examine 3,000 abstracts indexed under the term WORK. Experience has shown that in the majority of cases, any abstract dealing with sequences such as "thermionic work functions" is not necessarily indexed under all three terms, and consequently would not be obtainable from the above intersection. By some means, however, abstracts containing sequences as those cited above most often are indexed under that term in the sequence that happens to be a panacea term (in this case WORK). After a few failures with sophisticated intersection-type searches, the searcher is forced to rely on computer "dumps" rather than on intersections and complementations. This practice increases the output cost and reduces searching capacity and efficiency.

Analysis of more than 800 computer searches indicates that the number of accessions listed under each of the 250 panacea terms in the NASA MTV is increasing at the rate of 130 each month. This means that, on the average, every panacea term increases by 1,500 accessions per year, thereby losing its usefulness as a search term. The main reason for this phenomenon is a lack of preindexing term analysis. This means that present indexing criteria oblige the indexer to index under those terms appearing in the title and body of the abstract rather than terms describing the content of the abstract. Examples of this may be seen by considering computer searches using the following three terms: (1) LOW TEMPERATURE ENVIRONMENT, (2) NEGATIVE RESISTANCE DEVICE, and (3) MODULATION INDUCING RETRODIRECTIVE OPTICAL SYSTEM. The abstracts obtained by using Term 1 showed that EFFECTS rather than ENVIRONMENT would have more accurately indicated the subject matter. Only in a minority of the abstracts did the sequence "low temperature environment" appear in toto. In contrast, the majority of the abstracts contained only the sequence "low temperature" with the word "environment" appearing elsewhere in the body of the abstracts. It seems that the indexer, having a priori knowledge of the existing MTV term LOW TEMPERATURE ENVIRONMENT used this term as a further basis for indexing rather than differentiating between words modified by the sequence "low temperature."

The results obtained by using Term 2, NEGATIVE RESISTANCE DEVICE, indicate once more the indexer's tendency to ignore the content of the abstract and concentrate on previously established MTV term meanings. The use of the word "device" is both misleading and redundant as the majority of the abstracts did not deal with a device except in the extended sense of the word. The indexer considers a transistor as a "transistor device," a tunnel diode as a "tunnel diode device," ad infinitum. In this instance, indexing the abstracts under the term NEGATIVE RESISTANCE would have sufficed since this term completely characterized them.

The output due to Term 3 again proves that the indexer is unaffected by the content of the abstracts. More important, it shows that the terminology of a very small number of abstracts can be taken too seriously by the indexers. As a result, abstracts dealing with identical subject matter expressed in a slightly different terminology will not be indexed under the same MTV term. Six of the total of 11 abstracts that were indexed under Term 3 used the term either as their title or as part of their title. Of the remaining 5, one was completely misindexed, and the remaining 4 simply referred to the acronym MIROS which stands for Term 3. Since MIROS is a NASA-sponsored research project, the 10 abstracts reflected a particular terminology as well as a research area. The fact that no "open" literature (e.g., published articles in journals, books, etc.) was indexed under Term 3 can either mean that only NASA is interested in this area or that independent research along similar lines lags by at least three years. Both of these alternatives are untenable. Hence, it seems reasonable to assume that the absence of open literature on this subject, under this term, is due to difference in terminology.

The first step toward the elimination of panacea terms is to create indexer awareness of the problems they can create, some of which have been discussed above. The fundamental rule for eliminating panacea terms is that MTV entries should convey the content rather than the terminology of any abstract or collection of abstracts. In addition, indexing under single words whose meaning is a function of words they modify or words that modify them should not be allowed. Descriptive terms to be entered into MTV need not and should not necessarily be taken from the title of the abstract to be indexed. If a term allows a broad interpretation, it should modify or be modified by one or more terms in such a way that its ambiguity or broadness is reduced. If this is not feasible, the term, in spite of its appearance in an abstract, should not be entered into MTV. Every effort should be made to gear the content of an abstract to descriptors which are a part of the user's terminology, especially when multiple-modified terms are involved. Discretion should be used in picking word sequences or clusters from abstracts and entering them into MTV without first considering the user's terminology. For example, an abstract should not be indexed under *Intermolecular Bonding* simply because the cluster appears in it when it could be indexed under the equivalent and more popular term PRESSURE WELDING.

However, care should also be taken to avoid excessive modification of panacea terms. For example, consider the terms CROSSED FIELD AMPLIFIER, PULSE WIDTH MODULATION, and PLASMA ARC METAL SPRAYING. The first two terms are of the form AAN, while the third term has three adjectives—AAAN—modifying the noun. In essence, each successive adjective modification of a term represents the intersection of that adjective with the remaining adjectives and the noun; i.e., AN means $A \times N$, $AA'N$ means $A \times A' \times N$, and $AA'A''N$ means $A \times A' \times A'' \times N$. Thus, the original set AN shrinks (becomes more descriptive) as the number of adjectives increases. The indexer's ability to recognize that descriptive terms are useful, only to the degree that they do not limit meaning by overmodification, is essential to the proposed method for eliminating panacea terms. For example, the sets PLASMA METAL SPRAYING or PLASMA SPRAYING are no less descriptive than their mutual subset PLASMA ARC METAL SPRAYING, but are less restrictive.

The above proposed techniques, while helpful to the system, are, in reality, a trend toward increasing the number of generic levels for indexing. To a large degree, economics favors this trend provided it is carried out with discretion. If this is not the case, both input and output costs will rise. Implementation of this trend for achieving functional and economic success is dependent on feedback from the system's users. The following statement by Dr. Mortimer Taube is apropos to this discussion:

> One of the things we have noticed in designing systems is that one can post or index an item on a number of different generic levels. Indexing on a number of different generic levels in principle increases the input cost and reduces the output cost. A saving can be made on the input side by indexing on a single level and by making logical sums to provide answers to general questions on the output end. This, of course, is an expensive way to search and if your clients ask you for many questions which involve making logical sums, the cost of searching will go up. On the other hand, if you attempt to anticipate what your clients are going to want and post, that is, index, on all the generic levels you can think of, you will increase your input costs immeasurably.[1]

• Synonyms

Whenever a literature searcher makes a computer search on a specific topic, he uses those MTV terms which, as a function of his knowledge and reference sources, characterize the topic. If the search results prove inadequate he has no recourse but to make additional searches

[1] See B. E. Holm in Bibliography.

using different but related MTV terms. Very often, the searcher's lack of knowledge of those MTV terms which are synonymous creates the need for additional searching. The NASA system is constructed in such a way that, if each of two synonymous terms is placed on a separate search request sheet, the computer will consider them independently. Therefore, any abstract that has been indexed under both of the synonymous terms will appear in both search yields. Consequently the searcher or user is faced with the same abstract when he screens the search yields.

The other major problem connected with the use of synonyms can best be illustrated by the following argument: Let A and B be any two synonymous terms appearing in the MTV. Let X_1, X_2, X_3, X_4, X_5, X_6, X_7 be seven abstracts such that the following are true: Abstracts X_1, X_3, X_6, X_7 each contain both A and B; X_2 contains A but not B; and X_4, X_5 each contain B but not A. Then the indexer is free to choose one or more of the following alternatives for indexing these seven abstracts:

1. He may index according to the terminology of the abstract. This means that X_1, X_2, X_3, X_6, X_7 and X_1, X_3, X_4, X_5, X_6, X_7 would certainly be indexed under A and B, respectively.
2. He may index the abstracts using only the terms A and B, but should a choice exist, he is free to exercise his preference of one term over another. This means that each or all of Abstracts X_1, X_3, X_6, X_7 can be indexed under both A and B, under B only, or under A only.
3. He may index, whether a choice exists or not, under A, under B, or under both, simply because he knows a priori that A and B are synonyms but prefers one to the other. This means that he may index X_2 under B even though B does not appear in X_2.

Suppose that the indexing has been consistent with the alternatives referred to above. If the searcher dumps Term B, he will lose X_2 which contains only A. If he dumps Term A, he will most likely lose X_4 and X_5 since they contain B only. In either case some abstracts will be lost. If he dumps Term A and B separately, he will then be faced with the same abstract/s in both search results. In either case, the searcher loses some abstracts or is faced with redundancy. Since these problems are intimately connected with those discussed in the next two sections, suggestions for their solutions will be postponed until then.

● Hierarchy

The indexing of information is intimately connected with the ordering of terms as well as with their meaning. By far the most definitively elusive and difficult of all the information indexing concepts is that of term hierarchy. Closely related to hierarchy is the connectivity of terms. In general, hierarchy implies connectivity but not the converse. An easily digestible definition of term hierarchy is the graded ordering of terms in a vertical manner, such that the term with the broadest spectrum occupies the uppermost position. On the other hand, term connectivity, in general, possesses a nonvertical character. In this paper, term hierarchy, in addition to its vertical character, will be considered as a chain of sets. By this is meant that each term, except the uppermost or mother term, shall necessarily be a subset of the one immediately above it and a superset of the one immediately below it. The mother set will be a superset of every set in the chain. In the case of term connectivity, no such inclusive order will, in general, be assumed.

Before plunging into examples of some common types of hierarchy and connectivity, the method of representing them will be discussed. The old cliché about a picture being worth a thousand words has real meaning for the systematic indexing of technical literature. Pictorial or schematic representations of hierarchy and connectivity of terms are more easily read, assimilated, and retained. Linear graphs (straight lines) are the simplest form of pictorial connectives. Linear graphs are simple to construct, easy to follow, and possess dynamic form. On the other hand, lists of terms such as those found in technical thesauri are dull, rambling, and static and defy any attempt for retention. In addition to those shortcomings, lists of terms do not convey the hierarchy of a given family beyond the first order of magnitude.

The requirement if it exists, that an indexer must consult a technical thesaurus for indexing an abstract on a subject about which he has dubious knowledge, is proper but expensive. The consultation of a reference such as the *EJC Thesaurus* whenever there exists doubt regarding the use or nonuse of a term is time consuming as well. To further require that a 200-word abstract be indexed properly in less than 15 minutes under the above conditions is unrealistic. There are, however, other ways which may increase the accuracy and decrease the labor of the indexing task. Through the use of linear graphs, relationships and order regarding terms which represent varieties about a certain interest area may be detailed more effortlessly, forcibly, and quickly. It is envisioned that individual figures similar to those found in this paper may become much more fashionable than word lists. These figures, reproduced either on desk-size flip charts or on microfilm cards with an accompanying selector-reader, would essentially be a pictorial thesaurus. With such a thesaurus, a flip of the hand or the push of a button would diminish ambiguity and facilitate understanding.

The author at this point must acknowledge a major debt to the authors of the *Euratom Thesaurus*. Though limited to the display of term connectivity, this work served as impetus for the enlargement of the scope of pictorial representations to include term hierarchy and its consequences.

The terms JOINING, RADAR, and LASER have been chosen to demonstrate the concepts of hierarchy and connectivity as they should be used in creating a pictorial thesaurus.

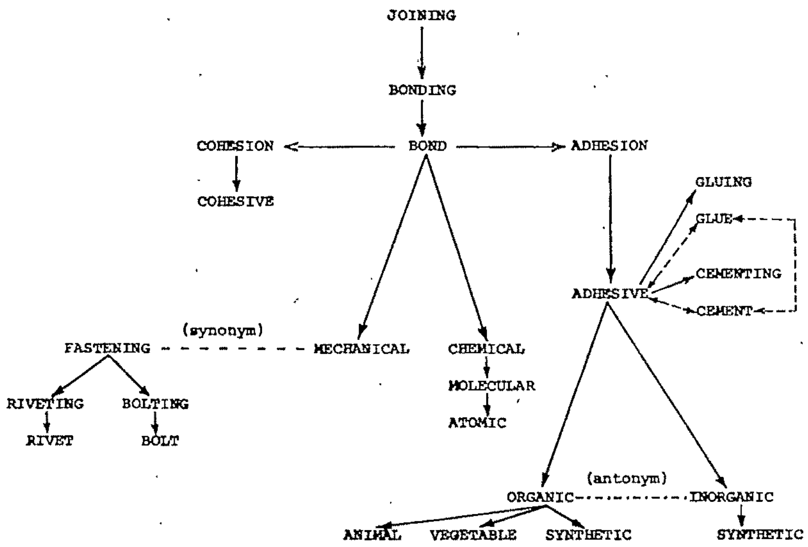


FIG. 1. Hierarchy of the term JOINING. (Synonyms designated by dashes [– – –]; antonyms designated by dash, dot [– • – •].)

Figs. 1 and 2 display the hierarchy of the term JOINING as defined by *Van Nostrand's Scientific Encyclopedia*. Together they represent no less than 2,988 words. Fig. 3 shows the hierarchy of the terms RADAR and LASER. Implicit in this attempt to show the relationship between terms is the intention of creating an awareness of different types of hierarchy such as those based on method, type, use, material, function, and adjective. The "function hierarchy" is best shown by the horizontal arrows between JOINING and WELDING, between SPOT and PROJECTION, and the double arrows between OXYGEN-HYDROGEN and ALUMINUM and HELIARC and MAGNESIUM. The "adjective hierarchy" is the successive modification of a term by adjectives. As an illustration, let RADAR be the mother term. Then TRACKING RADAR, RANGE TRACKING RADAR, RANGE TRACKING FIRE CONTROL RADAR, etc., constitute an adjective hierarchy. To further clarify the concept of function hierarchy, let f be the function called FORGING.

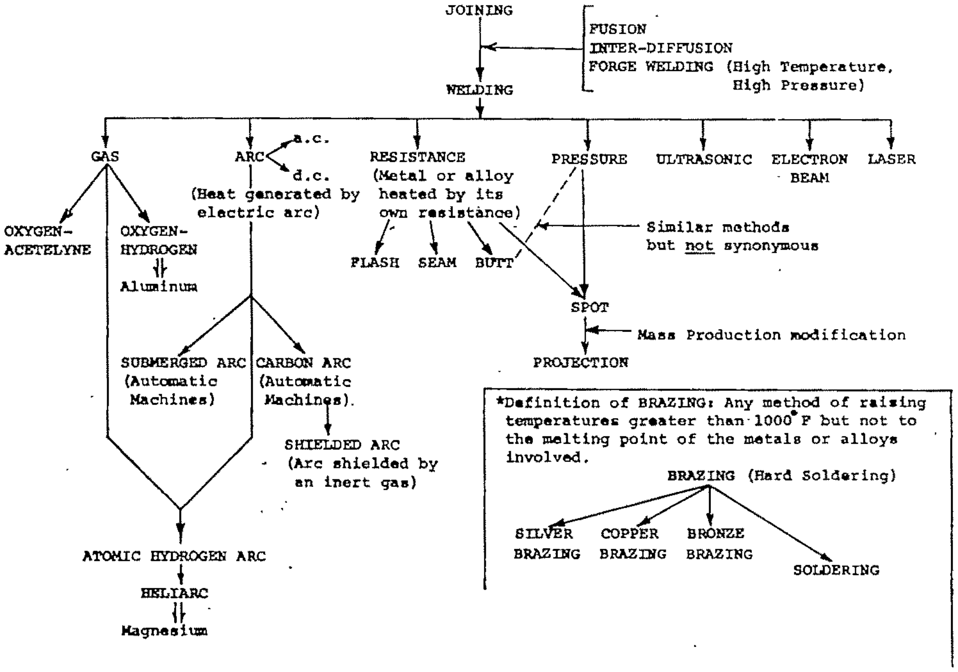

JOINING
FUSION
INTER-DIFFUSION
FORGE WELDING (High Temperature, High Pressure)
WELDING
GAS
ARC
a.c.
d.c.
(Heat generated by electric arc)
RESISTANCE
(Metal or alloy heated by its own resistance)
PRESSURE
ULTRASONIC
ELECTRON BEAM
LASER
OXYGEN-ACETELYNE
OXYGEN-HYDROGEN
Aluminum
FLASH
SEAM
BUTT
Similar methods but not synonymous
SPOT
Mass Production modification
PROJECTION
SUBMERGED ARC (Automatic Machines)
CARBON ARC (Automatic Machines).
SHIELDED ARC (Arc shielded by an inert gas)
*Definition of BRAZING: Any method of raising temperatures greater than 1000° F but not to the melting point of the metals or alloys involved.
BRAZING (Hard Soldering)
SILVER BRAZING
COPPER BRAZING
BRONZE BRAZING
SOLDERING
ATOMIC HYDROGEN ARC
HELIARC
Magnesium

FIG. 2. Hierarchy of the term JOINING.

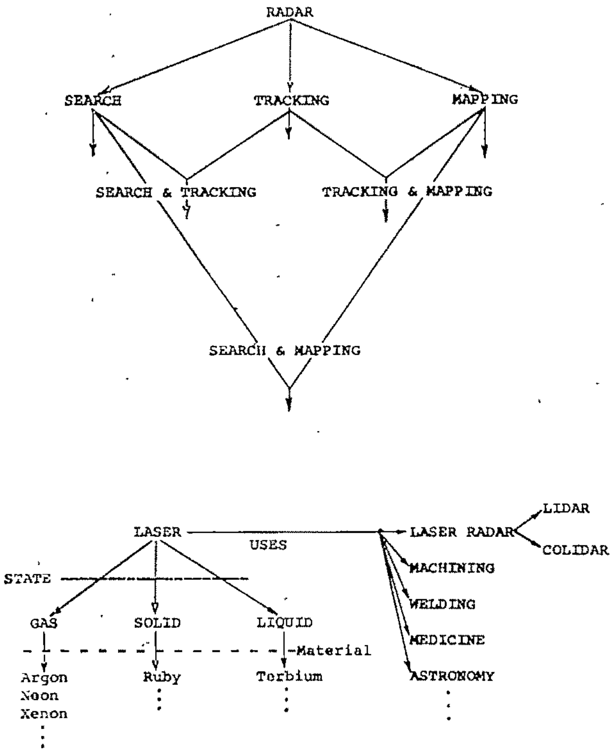
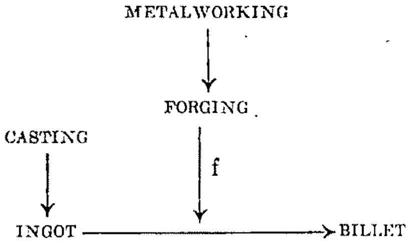


FIG. 3. Hierarchy of the terms RADAR and LASER.

Then the process of forging transforms an ingot into a billet or graphically:



From this graph it can be seen that a searcher, who was unfamiliar with this topic could obtain all information concerning the forging of billets by the intersection FORGING × BILLET rather than dumping FORGING or METALWORKING or CASTING.

• The Indexer's Six Infernos

The six infernos into which an indexer is likely to fall owe their existence to: ignorance, indecision, word frequency, terminology of the abstract, expediency, and too much or too little freedom for making terms eligible for MTV entry.

Ignorance is used to mean the indexer's lack of knowledge of a particular field about which an abstract relates. It is responsible for practically every indexing problem and is particularly responsible for the growth of panacea terms. The obvious solution is to employ indexers having broad scientific and technical backgrounds. Another solution to this problem might be to employ an indexer with broad experience to work closely with at most three less-experienced personnel. In this way, any questions regarding the indexing of an abstract could be answered immediately by the senior indexer, thus reducing the "ambiguity time" and increasing the accuracy.

The existence of too many choices of terms under which a given abstract may be indexed is in part responsible for *indecision* on the part of the indexer. Indecision is the prime cause of overindexing and can be minimized by discreet selection of MTV terms.

Word frequency is the practice of indexing an abstract under a term on the basis of its frequency of appearance in a given abstract. It is responsible for inaccurate and redundant indexing. While frequency of appearance is a necessary condition for indexing under a term, it is not a sufficient condition. An equally important consideration is the grammatical usage of the term in a given abstract; i.e., its usage as a noun, adjective, or verb. The indexer can seek to resolve this problem by taking into account the grammatical usage that a term implies. For example, a given abstract concerned with *radar ranging methods* may have "range," "ranging," "range radar," "to range," "ranging radar," and "radar ranging" in its title or body. To index this abstract under any term other than RADAR RANGING METHODS or RADAR RANGING would produce redundant indexing.

Terminology of the abstract means that whenever the indexer is confronted with variations in terminology of essentially similar concepts in different abstracts, he does not standardize the descriptive terms before proceeding to index these abstracts under appropriate MTV terms. This serious oversight is usually due to *Ignorance* and *Indecision*. A term's existence in 10 or 15 or even 50 abstracts should not be a prime criterion for MTV candidacy. A term's logical inexactitude should be considered prior to its entry into MTV. A term like SHORTENING is but one of many terms that must be judiciously inserted into MTV. For instance, it can mean "cooking oil," "Doppler shortening of electromagnetic waves," or even relativistic "shortening of swords" in the sense of FitzGerald. This problem of terminology will be rectified when the problems of ignorance and indecision have been resolved.

Expediency related to ignorance means that the indexer, in an effort to keep up some quota of indexing *m* abstracts per hour, tends to index certain abstracts in fields about which he has dubious knowledge under terms that do not reflect the content of the abstract but which have been previously established as MTV terms.

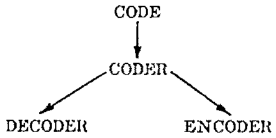
The following discussion will illustrate some of the pitfalls inherent in the six infernos referred to above. The first concerns the usage of too many terms of the same family as MTV entries. A representative family that exists in the NASA system is: CODE, CODER, CODING, DECODER, DECODING, ENCODER, ENCODING. Since information has to first be *coded* before it can be *encoded* or *decoded*, the term CODE emerges as the mother term.

If no hierarchy is defined for this family, then a literature searcher will, if he neglects to use all of these terms, lose some abstracts. This incompleteness in searching could, quite conceivably, mean that the search was useless. In addition to the loss of possibly pertinent abstracts, too many terms of the same family in MTV raise *both* the input and the output cost. As an illustration of the effect on output cost, consider the common accessions of the term CODE with each of the six terms of the family shown in Table 1.

To obtain the data for the table, single-term dumps were performed using each term. Next, the common accessions between each term and the largest term (CODE) were found manually. The first column of the table shows the logic code which could have been used to obtain the common accessions with the computer. The second column shows the percentage of common accessions of each term with CODE, relative to the number of accessions of each term. For example, the first row of the table means that out of 21 abstracts indexed under CODER, 12 of them were also indexed under CODE.

Fig. 4 is a schematic representation of all mutually common accessions of the six terms.

In contrast to using too many terms (of a given family), too few terms will produce the adverse effect of "forcing the indexing," thereby destroying the uniqueness of a descriptor. The two extremes conjure up a "law of spite" which may be partially abolished only by considering the logical connectivity of terms of a given family. For the family in question, the four terms CODE, CODER, DECODER, and ENCODER would amply suffice. Their hierarchy may be schematized as:



The second concerns the lack of rules for indexing under one or more terms having the same grammatical root. It is felt that distinctions should be made regarding the usage of *-ing, -or, -er, -ion* terms in MTV. Failure to distinguish between terms having these endings will result in an abstract being indexed haphazardly under some or all of the terms having the same grammatical root.

Some rules may be evolved by considering the premise that any thought, idea, or concept can be expressed in verb form. Consider the four tupple: DETECT, DETECTING, DETECTOR, DETECTION. The term DETECTING represents the verb form and hence the idea or concept, DETECTOR, being the necessary "tool" (machine, device . . .) for implementing the idea, and DETECTION representing the terminus of the chain. In symbolic logic form we have:

$$\left.\begin{matrix}\text{DETECT}\\ \text{DETECTING}\end{matrix}\right\}\Longrightarrow \text{DETECTOR}\Longrightarrow \text{DETECTION}$$

where the grammatical suffixes and the implication signs determine the genealogy.

TABLE 1.

CODE × CODER	12/21 = 57.1%
CODE × CODING	12/57 = 21.1%
CODE × DECODER	5/20 = 25%
CODE × DECODING	30/75 = 40%
CODE × ENCODER	4/20 = 20%
CODE × ENCODING	23/72 = 32%

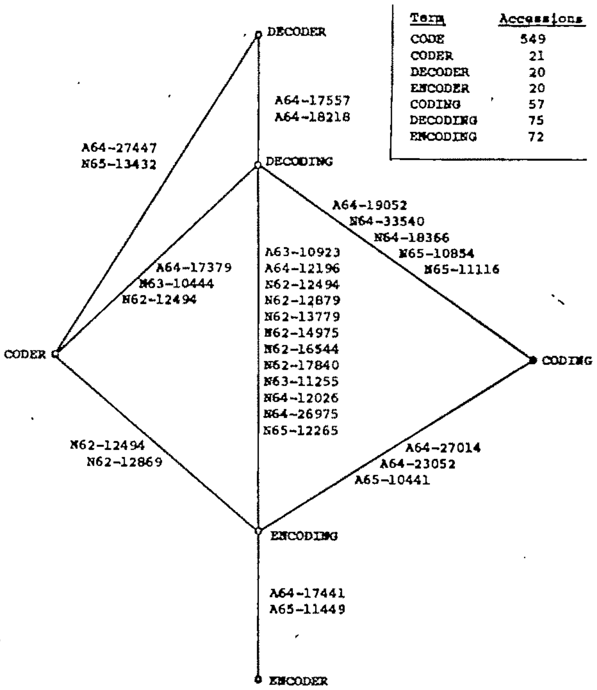


FIG. 4. A family of MTV terms and their common accession numbers. ("A" numbers refer to articles whose abstracts are found in the *International Aerospace Abstracts* [IAA]; "N" numbers refer to articles whose abstracts are found in the *Scientific and Technical Aerospace Reports* [STAR].)

• Negations and Antonyms

The NASA system allows searches to be formulated by using negation or complementation of terms, i.e., given any two MTV terms A and B, the information bank may be searched by either $A \times B^c$ or $B \times A^c$. Complementation may be used to advantage whenever it is known a priori what terms are to be negated. For example, suppose a searcher required information on the subject of *noise* excluding *radio noise, vibration noise*, and *stellar noise*. He could formulate his search request by $A \times (B+C+D)^c$ or $A \times B^c \times C^c \times D^c$, where A, B, C, and D stand for NOISE, RADIO, VIBRATION, and STELLAR, respectively.

The computer's ability to process negation of terms is essential to the overall flexibility and conciseness of the system. This ability, however, can be endowed only by the indexer. Negation of terms in English can be accomplished by prefixes such as *il, im, in, not, non,* and the suffix *less.* With the exception of those terms which can only be negated by the suffix *less,* many terms can be negated by more than one prefix. Negation by one or more of the tripples (un, not, non; in, not, non) is an example.

The absence of a uniform negation rule and the lack of correlation between synonyms, antonyms, and negations of MTV terms are a source of difficulty in working with the NASA system. In essence, the problem is that there exists no way by which a searcher, having found term "X" in MTV, can quickly and economically find all its synonyms or antonyms that appear in MTV. Should the searcher consult any number of thesauri for the synonyms or antonyms of a term "X," he has no guarantee that the author, abstracter, and indexer used the same sources. In cases where the assumption has been made that the same sources, (e.g., *EJC Thesaurus, Van Nostrand's Scientific Encyclopedia,* etc.) had been used, the search efforts proved inadequate. In addition, the trend toward single, double, and triple modification of terms can make the problem less tractable. For example, consider the terms NOISE ELIMINATOR, WHITE NOISE ELIMINATOR, and RANDOM NOISE REDUCTION. If the searcher were interested in the subject of *noise reduction* and this term did not appear per se in MTV, there would be no way of finding the above three terms because of the adjective camouflage. The only alternative would be to dump the term NOISE and screen 2,500 abstracts. This is a very expensive practice. The point is that indexing on different generic levels without a key to these levels is no better than indexing under a single level. A method which may overcome this difficulty would involve picking a term, creating its hierarchy, and making a list of all its synonyms and antonyms as they appear in MTV. This task is not as formidable as it might seem because, with the possible exception of panacea terms, these lists would be quite small.

Since every abstract reflects its author's or abstracter's preference of term negation, the indexer is forced, due to lack of a "negation rule," to follow suit. The result of this conforming will be the existence of families of MTV terms such as: UNSTABLE, INSTABLE, INSTABILITY, NONSTABLE, NONSTABILITY, NONSTABILIZED, and INFLAMMABLE, NONFLAMMABLE, and NONINFLAMMABLE. An extremely simple way of ridding MTV of multiple-negation forms is to employ a rule for grammatical negation. Such a rule might read: If any term is important enough to be inserted into MTV, and this term appears in any of the forms *il, im, in, not, non,* its negation is defined as *non,* plus the term. The case of *less* shall not be affected by this rule. Examples of this rule's duality are flammable, nonflammable; stable, nonstable; metal, nonmetal, etc.

At this point the two problems associated with synonyms will be discussed in terms of hierarchy, antonyms, and negations. The first problem concerns the possible appearance of one or more abstracts in the search yields of two or more synonymous terms whenever each term is placed on a separate search sheet. The second problem concerns the loss of those abstracts that were not indexed under both synonyms A and B. Essentially, both problems have a common solution. First, a hierarchy should be created where A, B, and all their synonyms are connected. Next, the synonyms should be ordered by "set inclusion"; i.e., if A, B, and C are three synonyms of graduated descriptive power, where A is the least descriptive and C the most descriptive, they will be ordered in such a way that C will be a subset of B and B a subset of A. Under this plan, the chance of losing any abstracts would be reduced by 50% since losses could only occur if the search were conducted with Term B.

Further, panacca terms such as NOISE should be considered the least descriptive in a family of terms which includes all adjective modifications as well as synonyms. Consequently all negations and antonyms should be considered relative to the term NONNOISE. Hence the terms QUIET, SILENT, SILENCE, NOISELESS, etc., shall be considered subsets of NONNOISE. This rule will facilitate searches involving the use of complementation.

Acronyms

The growth rate of acronyms in technical literature invokes the postulation of rules for indexing abstracts containing them.

The second edition of the *Space Age Dictionary* (SAD) contains approximately 140 acronyms and at least 200 symbols. It is estimated that 95% of the SAD entries appear in the technical literature contained in the NASA system. The majority of the symbols represent names of organizations or research projects rather than technical terms. Acronyms, however, do represent and are capable of contributing many technical terms to MTV. For instance, the acronyms RADAR, SONAR, LASER, MASER, and DORAN represent a set of at least 20 single technical terms. The panacea terms are denoted by an asterisk: RADIO,* DETECTOR, DETECTION,* DETECTING, RANGE,* RANGING, SOUND,* SOUNDER, SOUNDING, NAVIGATION,* NAVIGATING, LIGHT,* MICROWAVE,* AMPLIFIER,* AMPLIFICATION, STIMULATED, STIMULATION, EMISSION,* RADIATION,* and DOPPLER.

In addition, combinations of these terms may be formed, thus swelling the vocabulary without approaching in any way the descriptive power of the appropriate acronym. Examples of some combinations are RADIO RANGING, SOUND NAVIGATION, LIGHT RANGING, LIGHT AMPLIFIER, STIMULATED EMISSION, etc. It is obvious that, at this rate of combining and modifying terms, it will take upwards to 300 redundant or semiredundant terms to express what five acronyms can accomplish more concisely.

Without exception, acronyms are used as nouns or as adjectives. This fact leads to four possible cases for their appearance or for the appearance of their constituent terms in any abstract.

1. The acronym, modified by or modifying another term, appears in the title or in the body of the abstract.
2. The acronym, modified by or modifying another term, appears *only* in the body of the abstract.
3. The acronym, modified by or modifying or coexisting with some or all of its constituent terms, appears in the title or in the body of the abstract.
4. Some or all of the terms which form a particular acronym which does not appear in either the title or in the body of the abstract, appear in the title or in the body.

With regard to these four cases, the following indexing rules are given:

1. If, in a given abstract, the acronym is used as a *noun*, the abstract should be indexed *only* under the defining acronym. If no such acronym appears in the machine term vocabulary, it should be made to appear.
2. Under the premises set forth in Cases 1, 2, 3, or any combination of these, the abstract should *always* be indexed under the defining acronym. In certain cases the abstract may be indexed under *at most, one term* other than the acronym involved, *provided* that the other term reflects the central theme or main idea of the entire abstract and not simply the title.
3. If the acronym in either the title or the body of an abstract is modified by another acronym or by a term which itself describes either a *method* or a *concept*, then the abstract should be indexed under *both* acronyms or under the "noun-acronym" and under a *single additional* term best describing the method or concept involved. An example of one acronym modifying another is LASER RADAR. In this case, the *single additional* term could be COLIDAR. An example of a method or concept "acronym modifier" is DOPPLER RADAR.
4. An abstract containing an acronym must *not* be indexed under any or all of the constituent terms or their synonyms. For example, this rule forbids the indexing of an abstract containing the acronym RADAR under at least the following terms: RADIO, DETECTOR, DETECTION, RANGE, RANGING, all synonyms of these terms and all combinations of these terms.
5. If "acronym" is replaced by "symbol," the above rules still apply.

● Conclusion

This paper has demonstrated some of the more common indexing problems associated with any information system using a machine term vocabulary. Suggested methods for improving the system and providing feasible solutions to the stated problems have also been advanced. Though primitive, the partial solutions offered in this paper can upon further research, be refined and hence contribute toward the improvement of information systems.

Bibliography

1. *Thesaurus of Engineering Terms.* 1964. Engineers Joint Council, New York, N. Y.
2. *Euratom Thesaurus.* 1964. European Atomic Energy Community, Brussels, Belgium.
3. HOLM, B. E. 1962. Searching Strategies and Equipment. *Am. Doc.,* **13** (1).
4. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION. 1963. *Guide to Machine Searching and Retrieval of Information.* NASA, Washington, D. C.
5. *Van Nostrand's Scientific Encyclopedia.* 1958. 3rd ed. D. Van Nostrand, Princeton, N. J.

The User's Place in an Information System[1]

EDWIN B. PARKER [2]

Stanford University
Stanford, California

• Introduction

In our paper at the Airlie House Symposium (1) my colleague, Professor Paisley, and I characterized information-retrieval systems as receiver-controlled communication systems. We argued that information systems should be designed to maximize the amount of control by the receiver, and that, in general, the system should adapt to the receiver or user, rather than the user to the system.

We argued that the information needs and the information-seeking and information-processing behavior of scientists should be the subject of considerable psychological research. Such psychological research is required to provide adequate specifications for system designers and adequate criteria for the evaluation of systems.

• Future Systems

If we ignore all the constraints imposed by the limitations of existing technology, we can postulate a future information system with deep and flexible indexing, tailored to specific search requests. It might permit full text computer scanning of large numbers of documents according to criteria set by the user for that particular search. Let me illustrate the need for such flexible systems with a borrowed example (2). There are scientists who have the habit of piling up books and papers on their desks in a seemingly random fashion, yet know all the time how to find any given item. Should an assistant or a secretary bring apparent order to the desk, then the poor scientist may be unable to find anything. What is order to one person may be disorder to another, and vice versa.

The business of science itself can be said to be the creation or discovery of novel ways of viewing the environment such that hitherto unobserved forms of order come to light. This being so, it is not surprising that the categories used by a scientist in his current research often do not coincide with the categories of an inflexible index. Perhaps in some utopian future, all of us who consider ourselves scientists will be able to query the information system from computer consoles in our offices, according to indexing schemes prepared to fit the association patterns in our minds at the time of the query. Development of such a system would require as much research into the thought processes of scientists as it would research in systems design.

In the nearer future we will undoubtedly have to be content with indexing schemes prepared in advance of specific queries. Such relatively fixed indexes and classification schemes should, to be most useful, be based on prior research into the category systems and association patterns implicit in the minds of the potential users. Such research should investigate not only the categories and associations common to most of the potential user group, but also the range of idiosyncrasy in user information-needs. The deviant thinker may well be the more creative scientist.

• Three Caveats

There are three caveats that must be entered against this view of information retrieval as a receiver-controlled system and the concomitant suggestion that systems should adapt to users rather than users to systems.

One is the minor point that any information system can provide only what is available in the system, and to that extent is source controlled.

The second is that probably not all idiosyncratic search behavior should be accommodated. We may have to accept that a busy senior scientist with somewhat fixed patterns of information seeking may not be able or willing to change those habits. But we may want to teach younger or more flexible scientists more efficient search techniques.

The third caveat is that scientists do adapt to new communication systems. Just as the introduction of automobiles in our societies has profoundly influenced our social behavior, so new information technology will continue to influence our communication behavior. For example, it is usually clear at meetings such as this that jet air travel has considerably changed our use of both

[1] Presented at the ADI Symposium on Education for Information Science, Washington, D. C., October 10, 1965.
[2] Associate Professor of Communication, Stanford University.

informal and formal interpersonal channels for exchanging scientific information.

• User in System

Given these caveats it might be well to consider an alternate formulation. Another way of viewing the information-retrieval problem is to view it as a larger system in which the users are seen as part of the total information exchange system instead of outside it. In this larger system the problem becomes one of adjusting appropriate subsystems such that the flow of information to the scientist from all sources is in some way optimized.

One defect in this broader formulation is that it is extremely difficult, if not impossible, to specify criteria for the evaluation of such systems. Perhaps it might be possible to adequately specify "performance requirements" of engineers or scientists working on well-defined tasks. In such cases it should in principle be possible to evaluate the information flow within the system according to how well the engineer or scientist meets those performance requirements. I am less optimistic about such criteria for basic scientific research, however. Can we adequately measure scientific productivity? Scientists engaged in basic research implicitly set their own "performance requirements" by their behavior. This implicit setting of criteria within the system makes it difficult to visualize adequate explicit, external criteria for the evaluation of such systems.

Nevertheless, regardless of whether the user is viewed as inside or outside the information system, one conclusion is clear. That is the need for considerably more psychological research dealing with human factors in information-retrieval systems. There are several subareas of psychological research that seem particularly relevant. One is the physiological-perceptual research area that James G. Miller pointed at in his comments at the Airlie Symposium when he outlined the kinds of responses a human makes when confronted with an overload of information inputs. Another is the tradition of human factors research in man-machine systems. Another is the social psychological study of user needs and use of informal and formal communication channels. Still another is the area of cognitive theory and verbal behavior, which is highly relevant to the problem of developing classification schemes and association patterns that fit those of the users.

• Education

Those institutions engaged in teaching or research in information science, that have not already done so, might well consider hiring psychologists interested in information problems to supplement their present faculties and staffs. And since the demand for behavioral-science-trained information scientists is likely to exceed the supply for some time to come, some institutions might well consider introducing a PhD level program of education in this specialty.

It should be clear that information science can benefit from the detailed study of the most versatile information processing system ever developed or discovered — the human organism.

References

1. Paisley, W. J., and Parker, E. B. 1965. Information Retrieval as a Receiver-Controlled Communication System. In L. B. Heilprin, B. E. Markuson, and F. L. Goodman (Eds.), *Proceedings of the Symposium on Education for Information Science*. Pp. 23–31. Spartan Books, Washington, D. C.
2. Schafroth, M. R. 1960. The Concept of Temperature. In K. Messel (Ed.), *Selected Lectures in Modern Physics*. P. 268. Macmillan, London. Quoted in J. G. Miller, 1965, Living Systems: Basic Concepts, *Behav. Sci.*, 10: 201.

A Clearinghouse for Scientific and Technical Meetings: Organizational and Operational Problems

There has been much interest expressed in the formation of a clearinghouse for scientific and technical meetings. In spite of this interest, no clearinghouse has been formed. Whether or not there is a valid need for the existence of such a clearinghouse has not been proven, but in this paper it is assumed that the need exists. Discussion of the organizational and operational problems involved follows. These problems include: (a) definition of area and level of coverage of the subject matter of any given meeting; (b) definition of the geographic area from which the meeting will draw its attendance; (c) definition of the area of interest of the clearinghouse in the face of difficulties imposed by mission- or project-oriented meetings and other interdisciplinary meetings; (d) attaining comprehensive coverage in spite of the difficulties in obtaining inputs from the organizers of government classified and other closed meetings, as well as from the organizers of ad hoc meetings; and (e) the need to ensure that the organizers of meetings will make use of the clearinghouse. Solutions to these problems will not yield to a simple approach, but will be obtainable only on the basis of careful study of the structure and dynamics of the national and international scientific and technical meetings network.

HARRY BAUM

Technical Meetings Information Service
New Hartford, New York

The first complaint about meetings was almost undoubtedly made by Ptolemy at the time that Alexandria was a center of learning. It was at Alexandria that science began to develop along the line of "specialties" (1). It is, of course, specialization that has caused the information explosion of which the "meetings problem" is but one manifestation.

The most common complaint about meetings is that there are too many of them and that there is too much duplication and overlap. The usual cure offered is that a clearinghouse be set up that will enable the organizers of meetings to determine whether or not a similar meeting was already planned for a similar audience at a similar time (2, 3, 4, 5). In theory, a potential meeting sponsor, finding that such a conflict existed, would either call off his projected meeting or combine it with the conflicting meeting. While the idea is a valid one and is conceptually simple, it is by no means simple from the standpoint of organization and operation. It is these factors that I will discuss at some length. My discussion will be based on the experience I have gained in three years of planning and directing Technical Meetings Information Service (TMIS). I originally envisioned TMIS as a clearinghouse. It fell far short of that goal. The reasons will be apparent in the succeeding discussion.

Before getting on with the main discussion, however, I must clarify my position with regard to the complaints that exist about proliferation and overlap of meetings.

I neither agree nor disagree with these complaints. I believe that they are based, for the most part, on intuition rather than fact. A fair amount of contrary evidence does exist. For example, in the biomedical field a recent study by Orr and others (6) concluded that: "The large increase, over the past few decades, in the number of meetings at which biomedical research is reported has not exceeded the increase in the number of scientists engaged in such research and is a direct consequence of this growth in manpower . . . ." Similarly, in the field of psychology, a report by Compton (7) states: "Since membership [in the period from 1936 to 1961] has increased ninefold, an increase in programmed events could be anticipated, though the latter has, in fact, *not* kept pace with the gain in membership." A study by Pendray (2) of technical meetings in the flight sciences concludes: "After careful reading of the programs of eight major societies in the flight sciences, and detailed comparative study of the programs of four of these societies . . . we have come to the conclusion that there is relatively little, if any, overlapping of specific subject matter in the meetings of these societies."

I cite these contrary opinions, not to undermine the case for a clearinghouse, but rather to put what is often considered the primary reason for the existence of a clearinghouse in perspective. My own belief is that the need for a clearinghouse will become increasingly urgent in response to several clearly observable trends within the scientific and technical community. First, there is the growing interest in the applicability of the methods and results of other disciplines to the problems within a given discipline. Second, there is increasing emphasis on areas such as environmental science and engineering, and aerospace science and engineering that cut across almost the complete spectrum of science and engineering. Third, there is the rapid expansion of engineering and science into the public sector. I believe that these factors are causing a shift in the social structure of the entire scientific/technical community. As a result of that shift, the neat compartmentalization of science and technology that had formerly permitted us to keep track of any given discipline is being destroyed. The problem of keeping track of the movement of science and technology is increasing by an order of magnitude as a result. While there may have been little conflict among meetings in the past, this nice state of affairs cannot be expected to continue unless a mechanism is provided that can cope with the increasing order of complexity. The clearinghouse is such a mechanism.

But prevention of conflict among meetings should certainly not be considered the only function of the clearinghouse. A great amount of information is exchanged via meetings. They serve both as a formal and an informal medium for scientific and technical intercourse. The cost to the scientific and technical community of meetings is in the neighborhood of $1 billion per year. And, in terms of the sociodynamics of science and technology, a recent report on information use among scientists and engineers reveals that we may have seriously underestimated the importance of oral communication (8). A clearinghouse, in addition to serving to prevent conflict, could also serve as a central source of information about meetings for the use of the entire scientific/technical community. Perhaps most important, it could be the key information-gathering arm of a much-needed organization designed to investigate the sociodynamics of meetings with a view toward enhancing their efficiency as a component of the scientific and technical information exchange complex.

Having finished with my philosophical discursion, I will get down to the main purpose of this paper: the discussion of the organizational and operational difficulties involved in a clearinghouse. These difficulties include: (a) definition of area and level of coverage of the subject matter of any given meeting; (b) definition of the geographic area from which the meeting will draw its attendance; (c) definition of the area of interest of the clearinghouse in the face of difficulties imposed by mission- or project-oriented meetings and other interdisciplinary meetings; (d) attaining comprehensive coverage in spite of the difficulties in obtaining inputs from the organizers of government classified and other closed meetings, as well as from the organizers of ad hoc meetings; and (e) the need to ensure that the organizers of meetings will make use of the clearinghouse.

• Definition of Area and Level of Coverage

The first requirement for a clearinghouse is some method of classifying the technical content of the meetings. The difficulty of this task depends on the use for which the clearinghouse is intended. If it is intended merely as a conflict-resolution mechanism, then a simple coarse-grained subject-classification system similar to that recently adopted by the Committee on Scientific and Technical Information (COSATI) (9) would probably suffice. Such a classification, coordinated with information on date and geographic location, would serve as a coarse filter to screen out meetings that might be in conflict. Detailed examination would then indicate whether or not a true conflict does exist.

On the other hand, if the clearinghouse is to be used as a source of information on meetings for the use of the scientific and technical community as a whole, then a system capable of much more sophisticated discrimination is required. A system of subject indexing using tens of thousands of terms would probably be needed. Some of the broader meetings would probably require hundreds of terms to adequately describe their content. In addition to categorizing meetings by subject, it would probably be necessary to classify them by level of treatment of the subject. This type of classification has been discussed in some detail by Savage (3).

If the latter type of classification is to be used, the clearinghouse must have available to it, depending upon the degree of specialization of the meeting, anything ranging from merely the title of the meeting (if it is extremely specialized such as "Conference on the Conflicts concerning Two-Gas Atmospheres and Artificial Gravity for Space Flight") to the title of every paper to be presented and some indication of its degree of technical sophistication (for a very broad meeting, such as the Annual Meeting of the American Association for the Advancement of Science). Such a system would make possible a sophisticated information-retrieval system, and probably more important, a truly workable means for alerting the community to meetings of interest.

• Definition of the Geographic Area From Which a Meeting Will Draw Its Attendance

Meetings are geographically categorized as local, regional, national, or international. While the meaning of each of these words, per se, is reasonably unambiguous, they are not sufficiently precise to permit their use in a clearinghouse intended for conflict resolution. A meeting designated by the sponsor as "international" may draw only a handful of attendants from outside the country of origin. Societies holding a "national" meeting each year find that the number of attendants is strongly dependent on the location of the meeting. Some meetings designated as regional may draw attendants from all over the nation. (One such meeting is the "Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy.")

What is really needed to enable a clearinghouse to function as a conflict resolver is a description of the expected *distribution* of attendance. Obtaining such a statement is much more difficult than obtaining a simple statement of the overall area from which attendance will be drawn. I feel, however, that information on attendance distribution is essential to the proper functioning of the clearinghouse.

• Defining the Area of Interest of the Clearinghouse

Any workable clearinghouse must have a limited franchise. Certainly any attempt to cover all meetings, on all subjects, anywhere, would be more than unwieldy. My own guess of the number of meetings that would be involved by such an area of interest would place it in the order of hundreds of thousands, perhaps even millions, per year. Limiting the franchise to regional, national, and international meetings would probably reduce the number by one or two orders of magnitude, to the tens of thousands. Further limiting the franchise to scientific and technical meetings would reduce the number to the high thousands or the low tens of thousands. Even this number, however, might prove unwieldy if information in depth were to be made available. It is at this point, unfortunately, that further restriction of the franchise becomes difficult. The difficulty arises from the fact that the clearinghouse must deal with complete meetings, and that many meetings cut a very broad swath through the combined fields of science and technology. This problem is comparable to that which would exist if *Chemical Abstracts* or *Biological Abstracts* were required to deal with complete journals rather than with individual papers.

Perhaps an example will help make my point a bit clearer. Let us consider a discipline-oriented franchise — electronics. Electronics would have to include radio and radar astronomy; biomedical engineering and instrumentation; geosciences instrumentation; psychology embodied in human factors and display engineering; agriculture, forestry, and photogrammetry as involved in remote sensing of environment; meteorology as involved in weather radar; nuclear engineering and physics as involved in power generation; statistics of quality control and failure analysis; space sciences as involved in computers, navigational instruments, rendezvous control, electric propulsion, etc. I could, of course, expand the list, but it already encompasses, in addition to other branches of engineering, disciplines in the physical sciences, geological sciences, biological sciences, and even the behavioral sciences. That should illustrate my point well enough.

Choosing a mission-oriented franchise would complicate the problem at least as much. Consider, for example, the problem of choice that any of the following "missions" would entail: (a) environmental science and engineering, (b) communications and information, (c) space science and engineering.

My own tendency in deciding what to encompass in the *TMIS Technical Meetings Index* has been to admit all of science and technology. For a publication such as the Index, where comprehensiveness of coverage is a goal rather than a necessity, the choice has, so far, been practical. For a clearinghouse designed for conflict reduction, however, comprehensiveness within the area of the franchise becomes virtually a necessity, and the choice of that franchise requires careful analysis.

• Problems in Obtaining Inputs

The problem of obtaining comprehensive information on future meetings resolves itself into two phases. The first is learning of the existence of a meeting; the second is obtaining detailed information on the meeting after its existence is known.

The simplest part of the first problem is learning of the regularly recurring meetings of the established scientific and technical societies. One can rely on their being held every year at about the same time. The time and place of occurrence are usually set two or more years in advance. Even this problem, however, is complicated by the fairly rapid emergence of new societies.

Many societies, in addition to their regular meetings, will call a number of special, ad hoc, meetings during the year. These meetings are usually smaller than the "regular" meetings, and are organized on a shorter schedule, usually in the range of six months to one and a half years. Because of their irregularity, coupled with the shorter organizational time scale, these meetings are much more difficult to learn of than the regular meetings.

Of a still higher order of difficulty of acquisition, are the meetings that are organized *outside* the framework of the professional societies. Organizers of such meetings include government, educational institutions, laboratories, trade organizations, and industry. Some are regular meetings; others are ad hoc. Of the latter group, one can only know that such meetings will occur. It would require an extremely good intelligence system to enable their rapid acquisition.

Perhaps the most difficult meetings to learn of are the "classified" meetings held under the auspices of the armed forces. While the existence of these meetings, and their names, is not classified information, the armed forces are often reluctant to make their existence public knowledge for fear that attempts by unauthorized persons to attend will add to the administrative problems of the meetings. These meetings are considered to be extremely important ones in the industrial/military community and should, I feel, be included in any clearinghouse.

Learning of the existence of a meeting is not sufficient. One must also obtain information in detail. This implies that active cooperation of the organizers of meetings must be obtained. My experience has been that most organizers of unclassified meetings will willingly give information. Unfortunately, there are a handful of important organizations that are not eager to cooperate. Their general attitude can probably be summed up by the following paraphrases of correspondence and conversations I have had.

> We keep our members informed of our meetings by means of notices in our journal, and we aren't particularly interested in people who aren't members.

> Meetings are open only to members of the Society and their guests. There is no general solicitation of papers. I doubt that making information on our meetings public would serve a useful purpose, and it might have the awkward effect of leading all and sundry to assume that they can present papers at our meetings.

I'm afraid it would take a great deal of paper and reddish-purple ink to adequately convey my feelings about this sort of presumptuous snobbery. My real complaint — feelings aside — is that this attitude is irresponsible. It fails to acknowledge the need for information of the rest of the scientific and technical community, to say nothing of the need for information — and the right to it — of the general public.

The other problem area in regard to obtaining detailed information is the armed-forces supported classified meetings. This is not to say that the information is classified. Information about the meeting, including programs and abstracts is usually unclassified. Many of the organizers are willing to give information, but many are not. Again, the reason given for withholding details is that public disclosure of information might cause unauthorized people to attempt to attend or to obtain copies of the papers. While these reasons do have a certain amount of validity, they are also open to criticism on the grounds that they make information less available to workers in the field who should have access to it.

Three approaches may be taken by a clearinghouse to the overall problem of obtaining information on meetings. The first is that of passive detection. In this method, the clearinghouse examines existing published literature for information on future meetings. This method involves a great deal of effort and has the twin liabilities of the use of secondary sources of information: lack of timeliness and greater risk of inaccuracy than in the use of primary sources. The second method is that of active detection. Here, possible sponsors of meetings are canvassed on a regular basis. This method results in greater accuracy and may, depending on the frequency of the canvass, result in greater timeliness. Its disadvantage is that it is completely dependent on the clearinghouse's knowledge of who may be expected to sponsor meetings, and on the willingness of the sponsor to cooperate with the clearinghouse.

The third method involves the application of pressure in some manner so that the sponsor will consider it necessary to automatically provide the clearinghouse with the needed information.

For a clearinghouse that is *not* intended as a conflict-resolution mechanism, some combination of the first two methods would probably suffice. For a conflict-resolving mechanism, however, the third method would almost undoubtedly be required.

• The Need to Ensure that Organizers of Meetings Will Use the Clearinghouse

Any clearinghouse, to operate as a successful conflict-reduction mechanism, must have the active cooperation of the organizations among whom it is to reduce the conflict. While the utility of a clearinghouse can probably be shown on the basis of the overall "energy budget" of the scientific and technical information-transfer system, it is not a simple matter to convince any single sponsor of meetings that it is in his own best interest to surrender a portion of his autonomy to a regulative organization. And a properly functioning clearinghouse, even if not explicitly organized as a regulative body, would inescapably come to have that function implicitly. Most sponsors of meetings will probably agree that a clearinghouse would be a good idea; but as far as they are concerned, they know about the plans of other organizations of interest (or other organizations know of their plans) and they really don't need it. Whether or not such an attitude is justified, is certainly not known at this time. (As I noted earlier, my own guess is that it is justified now, but won't be for long.) Study will be required to determine whether or not this type of clearinghouse is needed. But if it is, then some method will have to be found to encourage active cooperation with the clearinghouse. One possibility for obtaining this cooperation would be in organizing the clearinghouse as part of an auditing agency, somewhat like the Audit Bureau of Circulation in the publishing industry. I have discussed such a proposal at some length in an earlier paper (10).

• Summary

I believe that a clearinghouse for information on meetings is needed now as a source of information on past meetings, and as an alerting service for the scientific and technical community. Furthermore, while a conflict-reduction mechanism may not be needed right now, the unifying tendency of the very broad interdisciplinary missions currently getting underway will make such a mechanism a necessity in the near future.

Because of the many complex problems that must be solved before such an agency can be established and put into operation, I believe that a clearinghouse should be started as soon as possible so that it can be operational when it is needed. This is not to say that I propose that we willy-nilly establish one now. After all, what I have discussed here is largely my own opinion. And even though it is mine, I don't propose to mistake it for fact. I would propose, however, that as soon as possible we undertake a program to uncover the facts and trends of the scientific/technical meetings complex as a major component of the overall information system within the sociological framework of the scientific and technical community.

References

1. SINGER, C. 1959. *A Short History of Scientific Ideas to 1900.* P. 63. Oxford University Press, New York and London.
2. *Technical Meetings in the Flight Sciences — A Report to the Daniel and Florence Guggenheim Foundation.* 1959. Pendray & Co., Bronxville, N. Y.
3. SAVAGE, C. F. 1962. Planning the Technical Meetings Complex. In *27th Midyear Conference of the American Petroleum Institute, Division of Refining.* San Francisco, Calif.
4. CRAWFORD, J. H. (Chr.) 1962. *Scientific and Technical Communication in the Government.* Report No. AD 299545. Department of Commerce, Office of Technical Services, Washington, D. C.
5. NATIONAL ACADEMY OF SCIENCES — NATIONAL RESEARCH COUNCIL, DIVISION OF MEDICAL SCIENCES. 1964. Communications Problems in Biomedical Research. *Federat. Proc.,* 23 (5, Part I): 1124–1125.
6. ORR, R. H., COYL, E. B., and LEEDS, A. A. 1964. Trends in Oral Communication among Biomedical Scientists: Meetings and Travel. *Federat. Proc.,* 23 (5, Part I).
7. COMPTON, B. E. 1966. A Look at Conventions and What They Accomplish. *Am. Psychol.,* 21: in press.
8. BERUL, L. H., ELLING, M. E., KARSON, A., SHAFRITZ, A. B., and SIEBER, H. 1965. *DOD User Needs Study, Phase I, Final Technical Report.* Vol. I. No. AD 615501. Department of Commerce, Federal Clearinghouse for Scientific and Technical Information, Washington, D. C.
9. FEDERAL COUNCIL FOR SCIENCE AND TECHNOLOGY. 1964. *COSATI Subject Category List.* 1st ed. No. PB 166877N. Department of Commerce, Federal Clearinghouse for Scientific and Technical Information, Washington, D. C.
10. BAUM, H. 1965. Auditing and Control of Scientific and Technical Meetings. In *Proceedings of the 12th Annual Convention of the Society of Technical Writers and Publishers.* STWP, Columbus, Ohio.

An Operating Model of a National Information System

The eventual configuration of any National Information System will require close coordination between the many existing indexing-abstracting services. Reduction of the overlap among these services is one of the important objectives of a national system. North American Aviation, Incorporated, has developed and implemented a system which solves some of the problems of linking services or networks together so as to combine maximum scope of information retrieval with minimum indexing and abstracting effort. The techniques used in this system are discussed and some proposals are presented showing how these systems techniques can be combined with other proposed methods for use in a National Information System.

J. L. EBERSOLE

North American Aviation, Inc.
El Segundo, California

Received 5 February 1965

• Introduction

The need for continuing improvement of our National Information System is recognized by government, industry, and our institutions whose missions contribute to this objective. As a result there are in existence and in formulation many systems which are designed to further this general objective, but which differ in the methods and organizational structure they see as necessary to improve the national information situation.

The extent of this effort is impressive as illustrated by the existence of approximately 300 independent abstracting and indexing services supporting scientific effort in the United States (1). These have a total publication of about 2 million citations a year. As more indexing and abstracting services are created, as the volume of knowledge increases, and as it becomes increasingly interdisciplinary in character there will be an increasing problem of overlap among these services. The major manifestation of this overlap consists of two or more organizations indexing and abstracting the same document.

For some time there has been emphasis on the cost of nonavailability of information, especially that caused by duplication of research effort. Additional emphasis needs to be given to the cost of the abstracting and indexing services which will provide information dissemination and availability of the required degree, since this can be expected to increase as the quantity of information increases. It is thus apparent that any national system, regardless of the degree of centralization or decentralization, must have as one of its objectives the reduction of abstracting and indexing costs caused by overlapping effort.

Studies have been made of this overlap on open literature. For example, it is estimated that approximately 35,000 journals are published throughout the world (1). Data on the coverage of these journals by all the services, both profession and project oriented, indicates that on the average each one is being covered nearly four times.

Eighteen of the 300 indexing-abstracting services are profession-oriented services. These 18 account for approximately one third of the 2 million yearly journal citations. It is estimated that their cost alone will increase from $7 million in 1961 to $25 million in 1971. An analysis of 17,000 journals covered by 11 of the 18 profession-oriented services in 1961 showed a 50% overlap in journal coverage among these 11 services alone.

Although these costs are dramatic, they are probably exceeded both absolutely and relatively by the duplication existing in the coverage of technical reports. This duplication cost extends not only through the community of indexing and abstracting services, but through the vast complex of libraries and information centers operating in hundreds of companies and governmental agencies. It thus affects not only the profits and costs of the 300 indexing-abstracting services and their subscribers, but impinges on the tax dollar and the general business profit dollar as well.

Several promising solutions to the duplication problem have been proposed. One of these is the use of "Modular Content Analyses," proposed by Lancaster and Herner (2). This approach consists of the preparation of a modular abstract which contains: (1) a citation; (2) an annotation; (3) three abstracts — indicative, informative, and critical; and (4) a set of modular index entries. This approach assumes that one organization would prepare "Modular Content Analyses" for documents or articles from specified sources, or for specified subject disciplines, and make these available to other indexing-abstracting services which, with minimum editorial effort, would use selected portions for their own publication. Realization of the potential benefits of this approach depends on an organizational and administrative system of cooperation which would effect agreements on: "who" would prepare Modular Content Analyses for specific subject areas or for specific journals; "who" would be sent copies of these; and, what costing procedure, if any, would be utilized.

A plan proposed by Robert Heller and Associates (1) recognizes the importance of the latter aspect. This plan would include the creation of "Organization X" which would be responsible for, in effect, reusing and diversifying the products (possibly modular content analyses) of the Profession Oriented Indexing-Abstracting Services. Its function would be to selectively disseminate these products to one or more of the 270 Project (or Mission) Oriented Services who would use them — again with minimum editorial effort on their part — in their Project Oriented secondary publications. The result would be a reduction in the duplication of indexing and abstracting effort between Project Oriented and Profession Oriented Services. The plan further recognizes the need for an administrative scheme of cooperative effort among the Profession Oriented Services to reduce the existent inter-service duplication.

An organizational scheme somewhat similar to the "Organization X" proposal is that of the "National IR Network Coordination Center" proposed by Jonker as part of a proposed national system of interlinking Information Retrieval Networks (3). In addition to serving as a standards originator and repository of indexes, search files, etc., this organization would search tapes from one or more networks for a given requester network. The product of the search would include citations and abstracts which the requesting network (or service) would then include in its own collection and/or publication with a minimum of editorial effort.

These proposals, and others, though differing slightly in approach, represent methods for improving our National Information System while at the same time reducing the costs inherent in the duplication which now exists in the system. Implementation of any of these approaches, no matter how promising they appear, is the major problem, however. The time and costs involved make it almost mandatory to "prove" the "workability" and feasibility of the potential solutions via pilot systems or by implementation in organizations or associations which have some of the characteristics of the national system and which would therefore serve as a model or microcosm of the National macrocosm.

North American Aviation, Inc., has implemented a system having many of the characteristics of a National system. This system's approach is complementary in concept, but more limited in application than some of the proposals discussed above. As such, it represents a study in the development and use of techniques by which some of the problems of a multivariate community of indexing-abstracting services have been solved. North American's subnational microcosm generates and utilizes information covering a broad range of products and subject disciplines and a growing amalgam of interdisciplinary tasks requiring access to both open and closed literature. These needs also result in variations of indexing approaches, especially those due to, in effect, a mixture of project and profession-oriented indexing.

The divisions of North American generate approximately 8,000 technical reports each year. In addition, the nine major libraries accession 52,000 external reports per year. These are available to users through 9 main libraries and 18 branch libraries. Duplication of indexing effort prior to the new system varied from 5 to 15% resulting in overlapping effort on as many as 6,000 reports each year. The objectives of the new system were to not only solve the "information explosion" problem but to eliminate the extra costs of indexing these "common holdings."

An exposition of this approach is only meaningful in the context of the North American Aviation Technical Information Processing System. Therefore, we will proceed to a general description of this system and then to the methods used to eliminate the effects of overlap in this milieu.

● North American Aviation Technical Information Processing System

Basically, this is a document retrieval system. Its output is similar to the secondary publications of many abstracting and indexing services. Specifically, these outputs are:

1. An Accessions Catalog. This contains a citation, a set of descriptive terms, and an abstract for each document received by North American Aviation Technical Information Centers and Libraries. Each division has a unique series of accession numbers which make up a separate section of the catalog.
2. A Permuted Descriptive Terms Index.
3. An Author Index.
4. A Source (Corporate Author) Index.
5. A Document Number Index.
6. A Contract Number Index.

In addition to the production of the Catalog and Indexes, the system also includes capabilities for SDI

(Selective Dissemination of Information) and Retrospective Search and Retrieval. The computer system for index and catalog production was implemented in the summer of 1964. The SDI and Retrospective Search subsystems were implemented in the summer of 1965.

The system is analogous to the "Union Catalog" approach in that the catalog and indexes combine bibliographic references for the holdings of nine geographically separate North American divisional information centers and central libraries: namely, Atomics International Division, located in Canoga Park, California; Autonetics Division, in Anaheim, California; Columbus Division, Columbus, Ohio; Los Angeles Division, Los Angeles, California; Rocketdyne Division, Canoga Park, California; McGregor Facility of Rocketdyne Division, McGregor, Texas; North American Science Center, Thousand Oaks, *California;* Space and Information Systems Division, Downey, California; and the Tulsa Facility of the Space and Information Systems Division, Tulsa, Oklahoma.

The North American system is divided into two levels of processing; a Division level system and a Corporate level system. The Division level system includes cataloging, indexing, and preparing abstracts (where abstracts are already in the document, these are used where possible, with modification as required). After these steps are performed, the information is translated into either punched cards or punched paper tape. The IBM 1401 computer is used to convert the paper tape or cards to magnetic tape, to perform certain audit and edit functions, and to create and update a Division Master File which contains all the accessions for a division. Each division normally operates its system on a weekly basis. The division level output includes a shelf list and an option for printing 3×5 cards. Once a month a tape containing the month's accessions is generated for input to the corporate level system. This is then sent via microwave or telephone line transmission to the Corporate Data Processing Center.

Fig. 1 is a block diagram representation of the Corporate Level System. The system utilizes an IBM 7010/1301, i.e., a combination of random access disks and magnetic tapes, for processing.

The first program in the corporate system merges the input from the nine division level systems; performs various audit, edit, and decoding functions; and generates the tape used for printing the Accessions Catalog. The second major program handles only the citation and descriptive terms which have been placed in document number sequence. This program detects common hold-

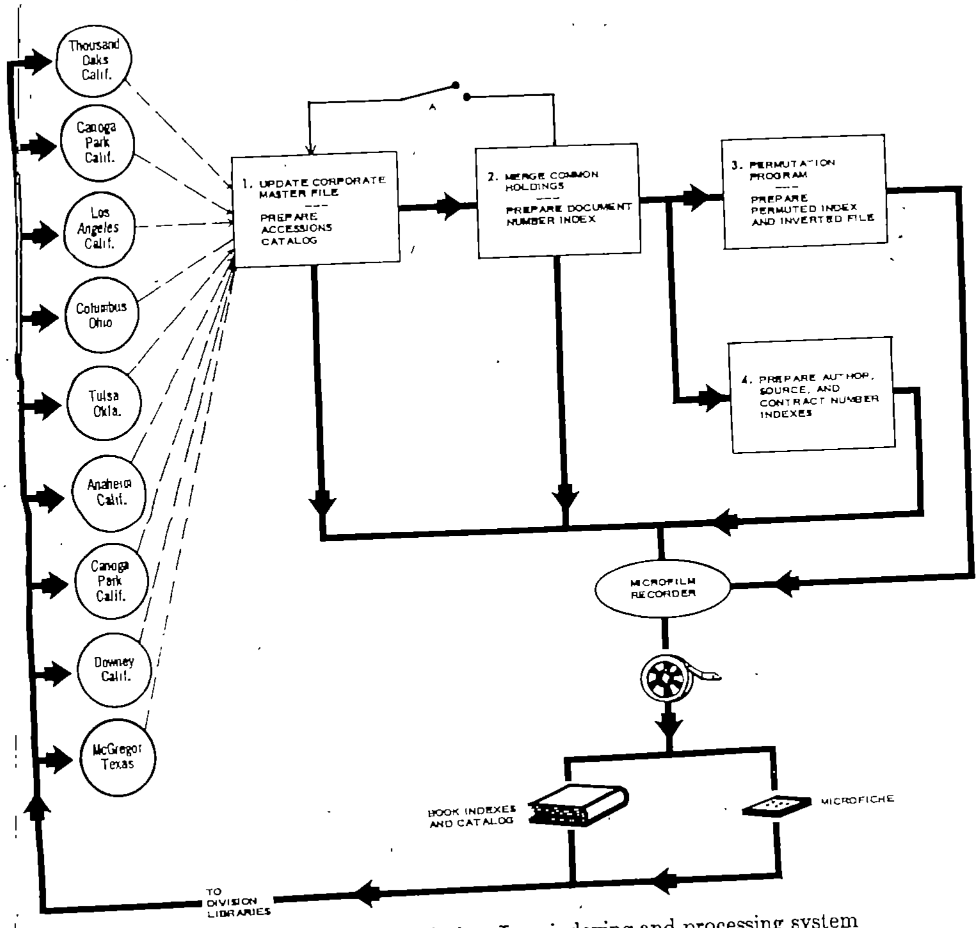


FIG. 1. North American Aviation, Inc., indexing and processing system general computer program flow.

ings, merges the records together, and prepares a tape which is used to print the Document Number Index. The third major program prepares the Permuted Descriptive Terms Index and creates inverted term files which are used for a term postings report and retrospective searches. The fourth major program prepares the Author, Source, and Contract Number Indexes. The output of all these programs is a "print tape" which is in SC 4020 (a microfilm recorder) language format. These tapes are run through the SC 4020 which prints the indexes and catalog on film at the rate of 6,000 lines per minute. This film is then used to directly create offset masters via Xerox copyflo. The published indexes and catalog are then printed using normal offset print procedures. Output options include the automatic preparation of microfiche from the film output so as to reduce the hard copy printing costs for those divisions which have the necessary reader/printers.

• Cooperative Indexing Scheme

In order to realize the cost savings provided by the capabilities of the system, it was necessary to establish a corollary administrative system. The administrative system is referred to as the "Cooperative Indexing" scheme.

To implement the cooperative indexing scheme, it is first necessary to assign indexing responsibility to a division. This is achieved by first determining which division receives all of the reports from a given source or receives a greater quantity and coverage of the output of a given source than any other division. The division is then contacted to verify if this, in fact, reflects a true discipline or mission interest. A division which meets these criteria for a given source can then be assigned the responsibility for indexing and preparing abstracts as necessary for all documents received from that source.

Sources for which the above determination can be made are designated as "Common Sources." The list of common sources and the division responsible for each one is coordinated periodically and disseminated to each division. When "nonresponsible" divisions receive documents from a common source, they prepare what is referred to as a skeletal input. This consists of that division's accession number, the source code, and the report number. The skeletal input may also include information that is unique to each division, such as branch library location, number of copies, etc. In order to make this approach feasible, a standard code for sources was developed. The provision for skeletal input of common holdings is one of the major reasons for the use of such a standard code. (The other reasons include achievement of consistency of sources so that a "clean" Source Index is possible and reduction in the punching and processing costs.)

Usage of the skeletal input is not restricted to the above situation. If a document from any source is already in the index — having been previously accessioned by another division — any other division receiving this document can then use the skeletal input. This can be easily determined by looking in the Source Index or Document Number Index.

• System Techniques and Methodology (4)

In addition to the general functions of master-file updating and producing an Accessions Catalog, the first major program of the NAA System performs certain other operations. One of these is to *generate a document* number record containing everything but the abstract for each bibliographic reference. During this process the accession number field, which is normally eight positions, is expanded by one position to include a "completeness" code reflecting the degree of completeness of the bibliographic reference. The program automatically assigns one of the following four "completeness" codes:

0=The input for this accession number contains no descriptive terms and no abstract. (This would be a skeletal input.)

1=The input for this accession number contains no abstract but does have descriptive terms. (This would occur when a "nonresponsible" division has a special interest in the document and may desire a different indexing approach for that document.)

2=The input for this accession number contains an abstract, but does not have descriptive terms. (This would not normally occur, but if it does, the user would be referred to the citation containing an abstract.)

3=The input for this accession number contains both an abstract and descriptive terms.

These document number records are then *sorted* in a modified sort program. Phase I of this *program expands* the document number *field into subfields* of 12 characters each with leading zeros, which, in effect, right justifies each number group in the document number. For this purpose, the program detects a numeric subfield in scanning from left to right as any group of numbers following a special character (such as a dash) or any group of numbers following an alphabetic character. Phases II and III of the modified sort program consist of normal sorting procedures.

In the second major step, the *"Document Number* Update Program," the records created above are used to update a document number master file, which *is in* sequence by document number in the natural sequence arrived at via the subfield technique. Normal updating procedures are used here except where a common holding situation is detected. This situation can consist of either two or more duplicate inputs (i.e., documents with the same report number and source code) in the monthly input, or one or more documents in the input which are the same as one or more documents already in the master file.

Fig. 2, representing variable length magnetic tape records, is an *illustration of the* merger that takes place in these situations. This example assumes that we have three divisions who have accessioned the same document. The division with the accession number L-084486 (the letter prefix designates the division) has prepared a complete input represented by Record 1. Division A in Record 2 has prepared a skeletal input and, in addition, has assigned descriptive terms. This situation could occur for several reasons, e.g., perhaps Division L has been assigned indexing responsibility for the particular document but Division A has a different indexing orientation due to a difference in mission and product. This may result in unique requirements for access to documents, which in turn require the use of certain terms (which may be division idioms or may be more specific terms resulting in greater indexing depth) which Division L does not normally use in indexing. Record 3 is a skeletal input from Division H. Because the system works with variable length records it is possible to add one or more descriptive terms and one or more accession numbers to any record.

The function of this part of the second program is to select the record to be saved — and added to — based on the highest completeness code. Since Record 1 has the highest completeness code it will be saved. In addition, the descriptive terms are compared as the records are merged. If the same term is found in all the records, the one contained in the record with the lower completeness code is deleted. If the terms do not match, the term from the latter record is added to the terms contained in the "save" record. The result of the merger is Record 4, which contains three accession numbers, the accession numbers from Records 2 and 3 having been added to Record 1.

The completeness code is used for an additional purpose in printing the indexes. When a user locates a potentially relevant document in any of the indexes, it is to his advantage to be able to refer to the bibliographic reference containing the most information about the document. To assist him in this, the accession number with the highest completeness code number is always preceded by an asterisk in the indexes.

Fig. 3 shows an actual example of the merger of a document accessioned by three divisions during the first months of system operation. Extracts from the Permuted Descriptive Terms Index, the Author Index, and the Document Number Index are also shown to illustrate how common holdings are printed in the indexes. The Permuted Descriptive Terms Index contains the set of descriptive terms and the document number and accession numbers for each document. The other indexes contain the title of the document plus the document and accession numbers. (The astute observer will note that our Indexing Training Course had not yet been completed when these inputs were prepared.) In this case, if you knew the report number, or if you were looking for documents by Mr. Gouse, or if you were searching for documents covering the subject term GAS FLOW, you would see references to the three accession numbers. Note that the "L" accession number would have the highest completeness code since it contains both an abstract and descriptive terms. Therefore, it is always preceded by an asterisk in the indexes. The completeness code for L-084486 would be 3. For A-000548 it would be 1, and for the skeletal input H-010309 it would be 0. This example also shows what happens when more than one division includes a title or an author in its entry. In this case the author and the title under the "L" accession number is saved. The "A" division's citation contains a contract number which is not contained in the "L" citation. Therefore, the merged record will also contain this number. The effect of the merger of descriptive terms can also be seen in the extract from the Permuted Descriptive Terms Index.

If a user from Division H came upon this document

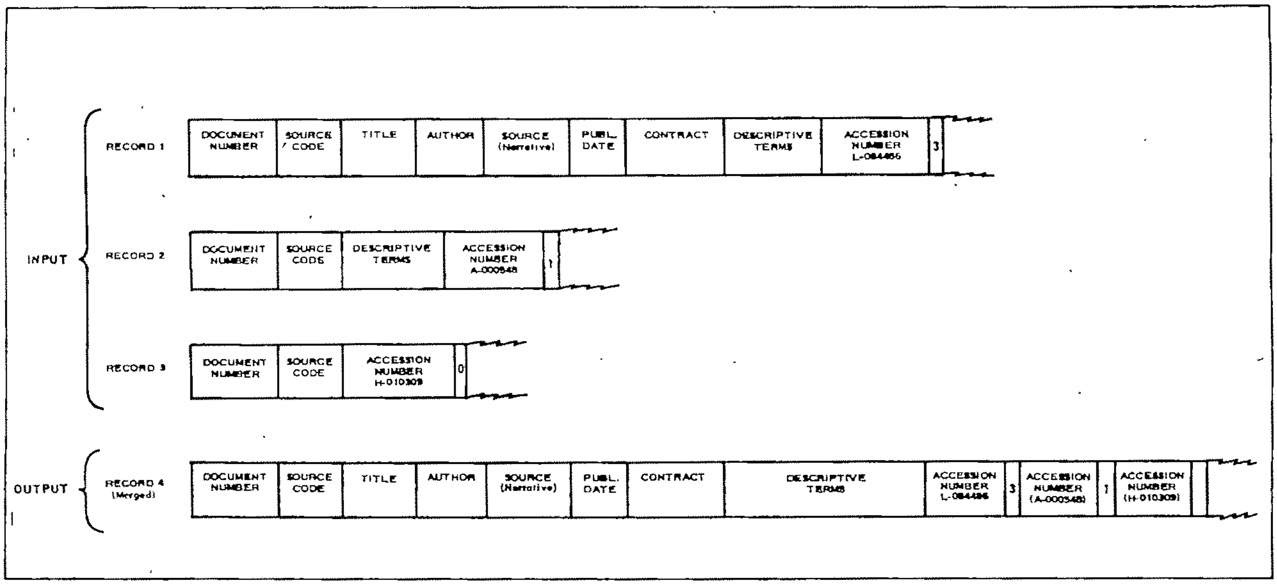
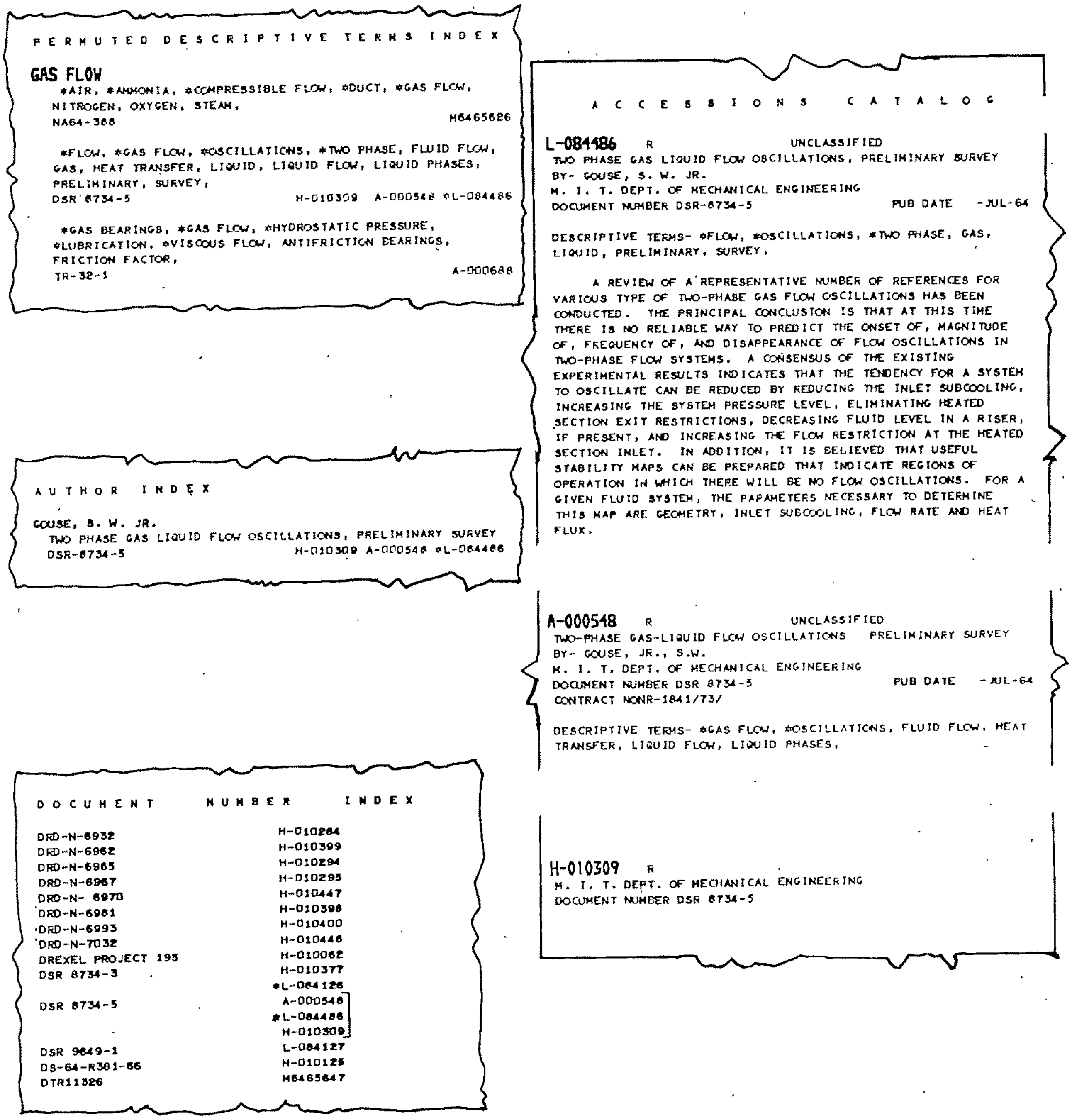


FIG. 2. Merger of common holdings.



```
P E R M U T E D   D E S C R I P T I V E   T E R M S   I N D E X

GAS FLOW
  *AIR, *AMMONIA, *COMPRESSIBLE FLOW, *DUCT, *GAS FLOW,
 NITROGEN, OXYGEN, STEAM,
 NA64-368                                            M6465626

  *FLOW, *GAS FLOW, *OSCILLATIONS, *TWO PHASE, FLUID FLOW,
 GAS, HEAT TRANSFER, LIQUID, LIQUID FLOW, LIQUID PHASES,
 PRELIMINARY, SURVEY,
 DSR'8734-5                       H-010309  A-000548 *L-084486

  *GAS BEARINGS, *GAS FLOW, *HYDROSTATIC PRESSURE,
 *LUBRICATION, *VISCOUS FLOW, ANTIFRICTION BEARINGS,
 FRICTION FACTOR,
 TR-32-1                                             A-000688
```

```
A U T H O R   I N D E X

GOUSE, S. W. JR.
  TWO PHASE GAS LIQUID FLOW OSCILLATIONS, PRELIMINARY SURVEY
  DSR-8734-5                    H-010309 A-000548 *L-084486
```

```
D O C U M E N T      N U M B E R      I N D E X

DRD-N-6932                    H-010284
DRD-N-6962                    H-010399
DRD-N-6965                    H-010294
DRD-N-6967                    H-010295
DRD-N- 6970                   H-010447
DRD-N-6981                    H-010398
DRD-N-6993                    H-010400
DRD-N-7032                    H-010446
DREXEL PROJECT 195            H-010062
DSR 8734-3                    H-010377
                             *L-084126
DSR 8734-5                     A-000548
                             *L-084486
                               H-010309
DSR 9649-1                     L-084127
DS-64-R381-66                  H-010125
DTR11326                       M6465647
```

```
A C C E S S I O N S   C A T A L O G

L-084486   R                     UNCLASSIFIED
TWO PHASE GAS LIQUID FLOW OSCILLATIONS, PRELIMINARY SURVEY
BY- GOUSE, S. W. JR.
M. I. T. DEPT. OF MECHANICAL ENGINEERING
DOCUMENT NUMBER DSR-8734-5                    PUB DATE   -JUL-64

DESCRIPTIVE TERMS- *FLOW, *OSCILLATIONS, *TWO PHASE, GAS,
LIQUID, PRELIMINARY, SURVEY,

     A REVIEW OF A REPRESENTATIVE NUMBER OF REFERENCES FOR
VARIOUS TYPE OF TWO-PHASE GAS FLOW OSCILLATIONS HAS BEEN
CONDUCTED.  THE PRINCIPAL CONCLUSION IS THAT AT THIS TIME
THERE IS NO RELIABLE WAY TO PREDICT THE ONSET OF, MAGNITUDE
OF, FREQUENCY OF, AND DISAPPEARANCE OF FLOW OSCILLATIONS IN
TWO-PHASE FLOW SYSTEMS.  A CONSENSUS OF THE EXISTING
EXPERIMENTAL RESULTS INDICATES THAT THE TENDENCY FOR A SYSTEM
TO OSCILLATE CAN BE REDUCED BY REDUCING THE INLET SUBCOOLING,
INCREASING THE SYSTEM PRESSURE LEVEL, ELIMINATING HEATED
SECTION EXIT RESTRICTIONS, DECREASING FLUID LEVEL IN A RISER,
IF PRESENT, AND INCREASING THE FLOW RESTRICTION AT THE HEATED
SECTION INLET.  IN ADDITION, IT IS BELIEVED THAT USEFUL
STABILITY MAPS CAN BE PREPARED THAT INDICATE REGIONS OF
OPERATION IN WHICH THERE WILL BE NO FLOW OSCILLATIONS.  FOR A
GIVEN FLUID SYSTEM, THE PAPAMETERS NECESSARY TO DETERMINE
THIS MAP ARE GEOMETRY, INLET SUBCOOLING, FLOW RATE AND HEAT
FLUX.

A-000548   R                     UNCLASSIFIED
TWO-PHASE GAS-LIQUID FLOW OSCILLATIONS   PRELIMINARY SURVEY
BY- GOUSE, JR., S.W.
M. I. T. DEPT. OF MECHANICAL ENGINEERING
DOCUMENT NUMBER DSR 8734-5                    PUB DATE   -JUL-64
CONTRACT NONR-1841/73/

DESCRIPTIVE TERMS- *GAS FLOW, *OSCILLATIONS, FLUID FLOW, HEAT
TRANSFER, LIQUID FLOW, LIQUID PHASES,

H-010309   R
M. I. T. DEPT. OF MECHANICAL ENGINEERING
DOCUMENT NUMBER DSR 8734-5
```

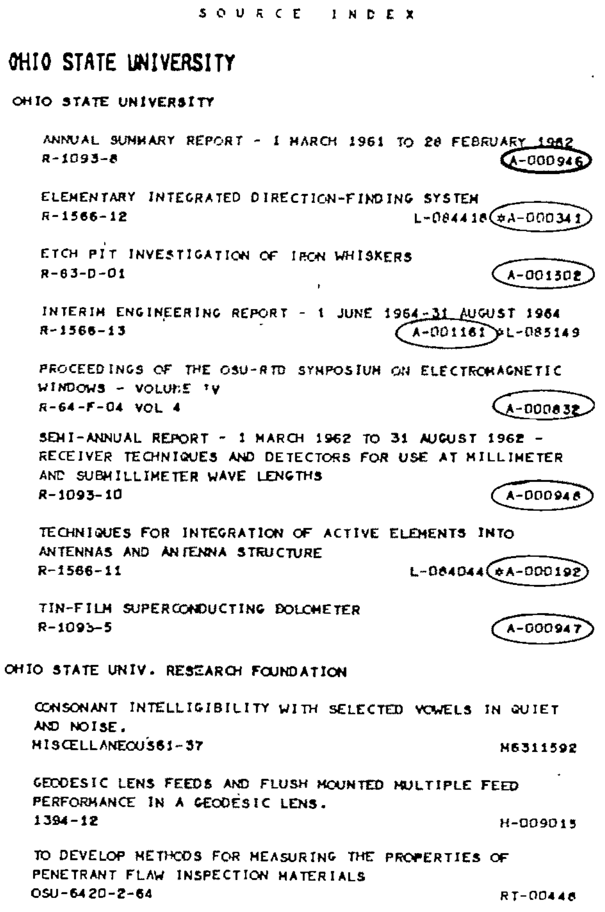
FIG. 3. Extracts from catalog and indexes.

via a search of any of the indexes he would turn to the "L" Accession Number bibliographic reference in the Accessions Catalog to find the most complete information. If, after reading the abstract, the document appeared to be relevant to his need, he would then check it out from his own library, requesting it under the "H" Accession Number. With this capability in the indexes, although the holdings of all divisions are retrievable, the quantity of documents ordered from other divisions is decreased significantly.

• Management Information Aspects

One of the first problems encountered in assigning indexing responsibility was that, in the previous non-integrated manual systems, it was impossible (without extensive expenditures of time) to determine the degree of overlap among divisions. It was also difficult for any given division to determine whether it received all of the documents published by a given source or whether it at least received more documents from a given source than any other division. Again, the common holdings technique solved this problem by providing information from which decisions on indexing responsibility could be made. In effect, this aspect of the system represents a facet of management information, at least for the purpose of the cooperative indexing scheme. The primary vehicle utilized for this purpose is the Source Index. After a large enough corpus of bibliographic references is built up for given sources, it is possible to determine answers to some of these questions. This is done by referring to each source in the Source Index. If every document listed under a given source carries the accession number of a given division, but only some of the documents also carry accession numbers for other divisions, it is possible to tentatively conclude that the former division should be assigned indexing responsibility for that Source. The conclusion is necessarily tentative at this point since it presumes the equating of coverage with subject competence or mission responsibility. The same approach can be used to decrease acquisition cost since the division which has the 100% coverage can in some cases perform the acquisition for other divisions.

The approach is illustrated by an extract from the Source Index shown in Fig. 4. Assuming the samples were of significant size, we can see that Division A has accessioned every document from Ohio State University. No tentative decision on indexing responsibility is possible for the Ohio State University Research Foundation since the three documents received during the period covered by this issue of the index were received by three different divisions.



```
                    S O U R C E   I N D E X

OHIO STATE UNIVERSITY

OHIO STATE UNIVERSITY

    ANNUAL SUMMARY REPORT - 1 MARCH 1961 TO 28 FEBRUARY 1962
    R-1093-8                                            A-000946

    ELEMENTARY INTEGRATED DIRECTION-FINDING SYSTEM
    R-1566-12                                  L-084418*A-000341

    ETCH PIT INVESTIGATION OF IRON WHISKERS
    R-63-D-01                                           A-001502

    INTERIM ENGINEERING REPORT - 1 JUNE 1964-31 AUGUST 1964
    R-1566-13                                 A-001161*L-085149

    PROCEEDINGS OF THE OSU-RTD SYMPOSIUM ON ELECTROMAGNETIC
    WINDOWS - VOLUME IV
    R-64-F-04 VOL 4                                     A-000832

    SEMI-ANNUAL REPORT - 1 MARCH 1962 TO 31 AUGUST 1962 -
    RECEIVER TECHNIQUES AND DETECTORS FOR USE AT MILLIMETER
    AND SUBMILLIMETER WAVE LENGTHS
    R-1093-10                                           A-000948

    TECHNIQUES FOR INTEGRATION OF ACTIVE ELEMENTS INTO
    ANTENNAS AND ANTENNA STRUCTURE
    R-1566-11                                  L-084044*A-000192

    TIN-FILM SUPERCONDUCTING BOLOMETER
    R-1093-5                                            A-000947

OHIO STATE UNIV. RESEARCH FOUNDATION

    CONSONANT INTELLIGIBILITY WITH SELECTED VOWELS IN QUIET
    AND NOISE.
    MISCELLANEOUS61-37                                  M6311592

    GEODESIC LENS FEEDS AND FLUSH MOUNTED MULTIPLE FEED
    PERFORMANCE IN A GEODESIC LENS.
    1394-12                                             H-009015

    TO DEVELOP METHODS FOR MEASURING THE PROPERTIES OF
    PENETRANT FLAW INSPECTION MATERIALS
    OSU-6420-2-64                                       RT-00446
```

FIG. 4. Example of use of source index to assign indexing responsibility.

Production of Mission or Discipline Oriented Indexes

One of the prime objectives of the North American information system is to make all of the technical information held by all North American libraries available to any North American engineer (except for those documents which may have restrictions on them). Because of this we have taken the union catalog approach. However, the same processing system with a slight modification could be used to provide selective indexes by Mission (or project) or by Discipline (or profession). This could be accomplished through the feedback loop created by closing Switch A in Fig. 1. This loop would consist of utilizing the common holdings technique to add the abstract, etc., from the "complete" input of a responsible indexing division to the skeletal inputs from other divisions. The result would be a catalog containing 100% complete bibliographic references for each division. The indexes would contain index entries for the holdings of each division plus indicating which documents were also held by other divisions. This can be achieved by a simple "select" program for the indexes which would select only index entries which contain that division's accession number series. Thus, if we assume that several of the divisions were Discipline or Profession oriented and that they indexed all documents in their field, and if we further assumed that other divisions were Mission or Project oriented and that they submitted skeletal inputs for documents which were of interest to them which had been indexed by the Discipline oriented divisions, the end result would be the utilization of the indexing prepared by subject specialists by the Project oriented divisions. Many other variations are possible with minor modifications in the system. For example, a given division might want its indexes to include every available document on a given subject. In that case, the entries which would be printed in those indexes would be selected by matching inputs from all divisions against each division's subject profile. Those which met the criterion would then be printed in that division's index. Within the purview of all the various proposals for national information systems and information networks there are many system requirements and objectives which could easily be handled by some variations of the North American system. These include the present capability to detect periodical "common holdings" (using the source code for the journal name and the volume, issue, and page numbers as a document number) and the capability to use the same type of merge technique for books (using author, title, and publication year for matching).

• Combining the Modular Content Analysis and the Common Holdings Approach

An extrapolation of a common holdings technique used in combination with modular content analyses would appear to be quite feasible. This possibility is suggested as one of the possible means of realizing the cost advantages inherent in modular content analyses.

The proposed modular content analyses contain three types of abstracts and two levels of index terms (the upper level could presumably be a subject category). If the modular content analyses were in machine readable form (e.g., paper tape or magnetic tape) and had certain standard internal computer assigned codes identifying each element of information, they could be sent to other abstracting-indexing services on an exchange or subscription basis. When another indexing service receives the tapes, it would have a set of established rules to apply in selecting the information it wishes to include in its own secondary publication or wishes to add to its own search files. The first test in this program would be to determine whether they wish to include the modular content analyses in their collection. This could be done in several ways. In some cases they might want to include every document issued by a given source. If so, this would be one of the acceptance criteria. Or, they might wish to include only information which met their own interest profile. If so, a normal search procedure would be used in determining this selection. After the initial decision was made, the program would then have to select the parts of the modular content analyses desired. Assuming that each of the three abstracts was identified in the machine readable media with a single digit identifying code, the program would merely select the code which had been assigned to the abstract it desired. If each higher level index term was also assigned a standard code (this would be simple if a subsumption scheme or a subject category scheme having a limited number of categories was used), the program could contain provisions for selecting specific terms subsumed under some of the headings, but possibly selecting only the more general terms for some subject areas. It would thus be possible for the using organization to select the depth of indexing they desired for each document. Such elements as author, title, etc., would be automatically selected whenever the initial criterion was met.

• Use of the Technique to Tap Files from Other Indexing-Abstracting Services

Since North American Aviation is a space contractor, the NASA master files (representing the content of the STAR and IAA indexes) are available for use. In addition, tape subscriptions are now becoming available from other indexing-abstracting services. Such tapes will be used for searches but, in addition, they will in some cases be used to reduce indexing costs even further by treating NASA or any other service involved as a "responsible division," for documents which they also index. In most cases, this will have only limited application since those divisions which obtain microfiche for most or all of the STAR holdings will use the STAR index. There are two situations, however, where the common holdings technique will be utilized. The first is applicable to smaller divisions who have in their holdings only a small quantity of reports which are indexed in STAR. A program has been designed which will allow a division in this situation to submit a skeletal input for any document which is, or will be in STAR. The program will then select the citation for that document from the STAR tapes, resulting in a reduction in the indexing effort required. The second use will be where information on a document is required in the division files so that it can be used as part of their circulation control and/or inventory system. By use of the skeletal input, a division can economically augment its files so that information on all of their holdings are available for those and other types of library administrative control systems.

• Use of the Common Holdings Technique in Future Systems

Many of the proposals for future information systems include the use of terminal devices in libraries or information centers which will be used for remote inquiry of mass-storage units. Developments in the software field indicate that in the near future we can also expect to have workable and economical automatic indexing systems. Assuming that remote inquiry devices were utilized, the common holdings technique can still be used to advantage. For example, when a North American division received a document it would first query the mass-storage unit to see if the document had already been indexed by someone else. This would be done by typing in a skeletal input. If the system "answered back" that the document was already indexed, the inquiring division would merely add its accession number as an input to the mass-storage index file. Thus, the same basic approach will continue to be applicable to decentralized information systems as these systems evolve by adopting new techniques and equipment.

References

1. HELLER, R., and ASSOCIATES. 1963. *A National Plan for Science Abstracting and Indexing Services.* National Federation of Science Abstracting and Indexing Services, Washington, D. C.
2. LANCASTER, F. W., and HERNER, S. 1964. Modular Content Analyses. *Proc. Am. Doc. Inst.*, **1**: 403–405.
3. JONKER, F. 1964. *A Model Information Retrieval Network for Government, Science and Industry.* Contract AF 49-(638)1209. Available from OTS, AD-600221.
4. WORTH, J. M. 1964. *System 2480 Data Processing Systems Manual.* North American Aviation, Inc., El Segundo, California.

Brief Communication

The SLIC Index

The principal weakness of conventional indexes is that, though there need be no limit on the complexity of the subject designations used, there can be great difficulty in matching those designations when a search is made. The indexer may assign to a document every term which the searcher can conceivably use, but the search requirement is likely to be specified by only *some* of those terms. Because the indexer has used terms with which the searcher is not concerned, the entries relevant to the search are scattered throughout the file. To take a simple example, the searcher requiring information on "the stability of single-engined aircraft" may be faced with the problem of extracting relevant items from a file consisting of entries such as:

AIRCRAFT, EXECUTIVE, SINGLE-ENGINED
— Economics
AIRCRAFT, EXECUTIVE, SINGLE-ENGINED
—Stability
BOMBERS, JET PROPELLED, SINGLE-ENGINED
— Control
BOMBERS, JET PROPELLED, SINGLE-ENGINED
—Landing—Stability
BOMBERS, JET PROPELLED, SINGLE-ENGINED
— Maneuvers
BOMBERS, JET PROPELLED, SINGLE-ENGINED
— Stability
BOMBERS, CARRIER-BORNE, SINGLE-ENGINED
— Control
BOMBERS, CARRIER-BORNE, SINGLE-ENGINED
— Design
BOMBERS, FOUR-ENGINED — Take off — Stability
FIGHTERS, SINGLE-ENGINED — Stability
FIGHTERS, SINGLE-ENGINED — Vulnerability

Concept coordination systems solve this problem completely. The searcher using an optical-coincidence card system would stack the cards chosen for his search, and the fact that other terms had been used by the indexer would not affect the success of the search in any way.

It is generally believed that the solution to the problem as it affects conventional indexes is to permute all the terms in a given heading. (The word "permutation" is used in its mathematical sense here. KWIC and like indexes are not permuted indexes, but are "rotated" or "cycled." More properly called permuted indexes are those using the Universal Decimal Classification, where class numbers are arranged in different orders, using the colon.) Permutation undoubtedly provides for retrieval by citing any combination of terms in any order, for the multiplicity of entries must necessarily cater for every possible approach. It is, however, not only extravagant but quite unnecessary. Our requirement is to provide for retrieval when any *combination* of terms is used by a searcher, and it is combinations in the mathematical sense, and not permutations, with which we are concerned. We need to provide in our visible index every combination of terms from the total number of terms assigned to a document by the indexer. If n is the number of terms assigned by the indexer, we must provide entries consisting of every combination of 1 from n, every combination of 2 from n, every combination of 3 from n . . . every combination of n from n (i.e., ${}_nC_1 + {}_nC_2 + {}_nC_3 \ldots + {}_nC_n$). This can be expressed simply as $2^n - 1$. We have not taken into account the question of the order in which the terms are to be cited, but in fact the index can function on the basis of combinations *only* if alternative orders are rejected (which means rejecting permutation) and a fixed order of terms in each heading is adopted. The obvious order for an index using ordinary language terms is alphabetical.

Let us suppose that an indexer assigns four terms — A, B, C and D — to a document and that we use alphabetical order as our standard order of terms in each heading. The entries which he needs to make in order to provide the facility for concept coordination in an ordinary visible file are as follows:

1. A
2. AB
3. ABC
4. ABCD
5. ABD
6. AC
7. ACD
8. AD
9. B
10. BC
11. BCD
12. BD
13. C
14. CD
15. D

If the same terms were used in an optical coincidence card system and a search was made for, say, the subject represented by the terms B and C this document would be retrieved though it has the additional terms A and D. It is clear, then, that a search for BC in a visible file which contains the entries listed above would be satisfied by the one entry BCD (No. 11) and that the entry BC (No. 10) is superfluous, for any entry consisting of *or beginning with* the sought terms is relevant to the search. It is therefore possible to reduce the number of entries required in a visible file still further, for all entries which form the beginnings of larger entries can be dispensed with. Thus, in the list above, Entries 1, 2, 3, 6, 9, 10, and 13 can be deleted and the remaining entries are as follows:

1. ABCD
2. ABD
3. ACD
4. AD
5. BCD
6. BD
7. CD
8. D

This is the absolute minimum number of entries required and the total is now reduced from $2^n - 1$ to $2^{(n-1)}$, i.e., from 15 to 8 in this case. Table 1 shows a comparison of the num-

TABLE 1.

No. of terms assigned to document by indexer	Permutations ($n!$)	All combinations $2^n - 1$	Selected combinations $2^{(n-1)}$
2	2	3	2
3	6	7	4
4	24	15	8
5	120	31	16
6	720	63	32
7	5,040	127	64
8	40,320	255	128
9	362,880	511	256
10	3,628,800	1,023	512

ber of entries produced by the three methods: permutation, the $2^n - 1$ formula, and the $2^{(n-1)}$ formula.

It is clear that the number of entries required in a visible file to give this facility is still too high to make conventional indexing on the usual card index principle a viable proposition if the average number of terms used for a document is high, for the labor of producing and filing so many entries would be excessive. The index which has been produced at ICI Fibres Limited does in fact use the formula $2^{(n-1)}$ but the principle of using an accession number to identify documents has been used, as with an optical coincidence card system, and the entire process of generating the appropriate entries from the group of terms used by the indexer, and the printing of the index, has been assigned to an IBM 1401 computer. The maximum number of terms which can be assigned to specify a subject has been set at 5, which means that such a subject receives 16 entries in the index. This limitation on the number of terms is the one disadvantage of the system as compared with an optical coincidence card or any other type of coordinate system, but any number of 5-term sets can, of course, be assigned to a document. Within these limits, concept coordination is as positive as with any other system. Because the index lists combinations which are a selection from the total number of possible combinations, it has been dubbed "The SLIC Index" — Selective Listing In Combination. The following is a description of the practical production of the index.

Documents are entered in an accessions register in order to assign to each a unique number, as for an optical coincidence card system. The same vocabulary is used as is used with the optical coincidence card index (ICI Fibres has such a card system using the French "Selecto" equipment), except that no term of more than 14 characters is acceptable because of the limited field sizes used on the IBM cards. The indexer assigns to a document a group of terms, not exceeding five in number, and these terms he enters on an indexing slip in alphabetical order, together with the accession number of the document, e.g.:

AIR: COOLING: CRYSTALLIZING:
SPHERULITES: TEMPERATURE 1394

This information is punched into a standard IBM 80-column card which is divided into 5 fields of 15 columns each and one of 5 columns. The 15-column fields each take a term and the 5-column field takes the accession number.

These "primary" punched cards are fed to the IBM 1401 computer which is programed to reproduce the "derived" cards forming the additional entries according to the prescribed formula, and this it will do whether a primary card has been given 5, 4, 3, or 2 terms by the indexer. The entire file of cards is sorted into strict alphabetical order and fed again to the computer which prints the index in the designed format. The computer is programed to print a given combination of terms once only and to subsume to that combination all the document numbers which are relevant. A sample of the finished index is appended (Fig. 1).

When a search is made, the searcher selects from the vocabulary those terms which he feels best describe his subject, as though he were going to use a set of optical coincidence cards. He jots down the terms in alphabetical order and consults the index at the point where this combination of terms appears. He takes into account all entries which *comprise or begin with* his chosen group of terms, notes the accession numbers, and consults the accessions register to identify the documents. He may find cases of the same number appearing under several headings, but once he has encountered the number he ignores all other citations.

Acknowledgment

This paper, without the appendix, was presented at the 38th Annual Conference of Aslib, 28 September to 1 October 1964, at the University of Exeter, with the permission of ICI Fibres Limited and Messrs. André Deutsch Limited who published the author's book, *Some Fundamentals of Information Retrieval,* which includes the substance of the paper, in March 1965. Aslib's permission to reproduce the paper here is gratefully acknowledged.

John R. Sharp
ICI Fibres Limited
Pontypool, Great Britain

Appendix

Simple Method of Determining Combinations To Be Used As Entries

The terms assigned by the indexer (a maximum of 5 in our case, which we will call A B C D E) are arranged in alphabetical order. The first term is then set down alone, and thereafter the following pair of rules is applied repeatedly until all the terms have been used:

1. Add the next term to all existing combinations.
2. Repeat all the combinations so formed, deleting the penultimate term in each case.

Setting the first term down alone, we have:

A

Applying Rule 1 we convert this to a new combination by adding the next available term which is B:

A B

Applying Rule 2 we remove the penultimate term, which is A, to form a new combination, which is B alone. We now have combinations:

A B
B

Rule 1 gives A B C and B C and Rule 2 adds the new combinations A C and C. Total combinations are now:

A B C
B C
A C
C

Rule 1 gives A B C D. B C D. A C D, and C D; Rule 2 adds A B D, B D, A D, and D. Total combinations are now:

A B C D
B C D
A C D
C D
A B D
B D
A D
D

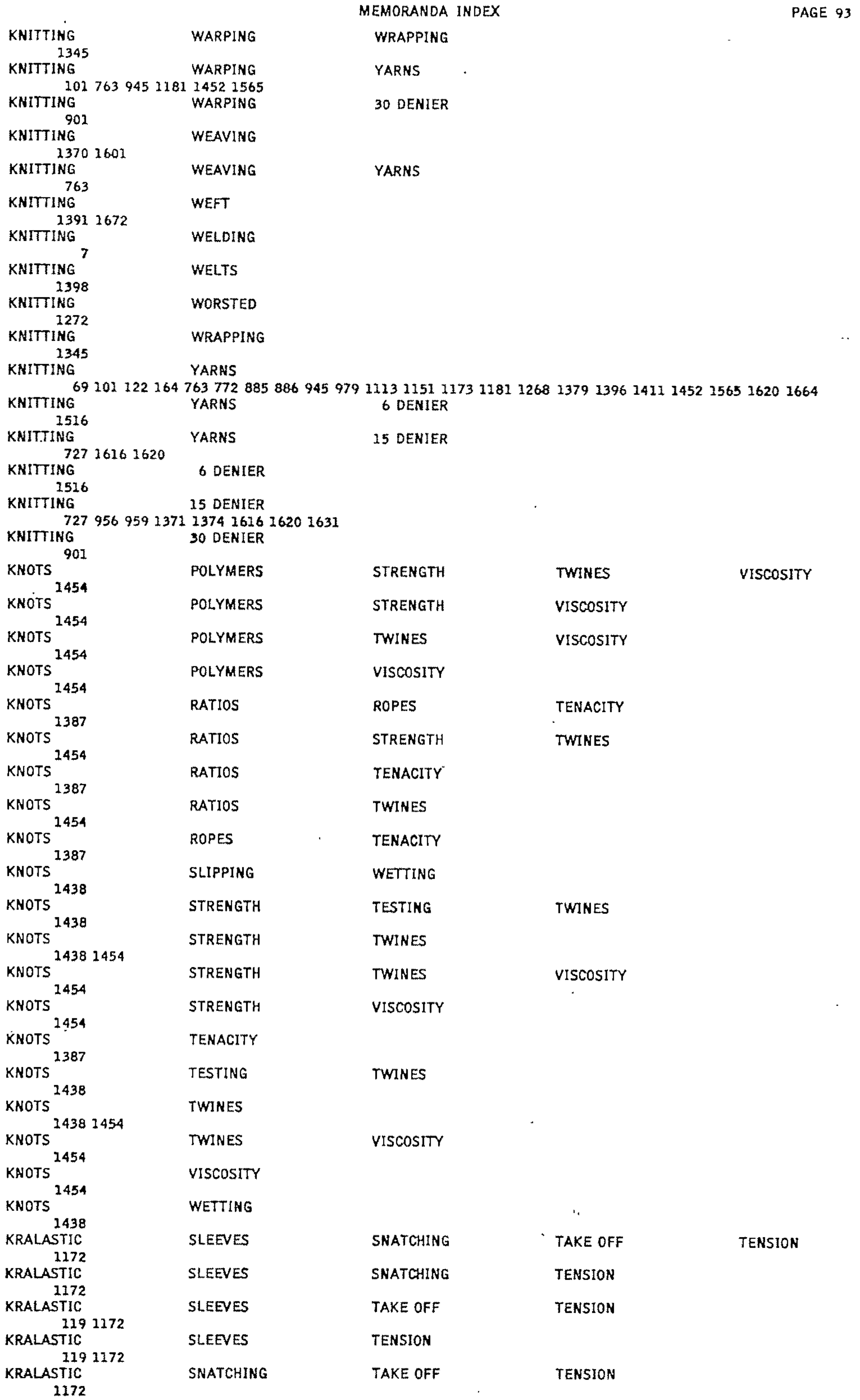
Rule 1 gives A B C D E, B C D E, A C D E, C D E, A B D E, B D E, A D E, and D E; Rule 2 adds A B C E. B C E, A C E, C E, A B E, B E, A E, and E. The final list of combinations is:

A B C D E
B C D E
A C D E
C D E
A B D E
B D E
A D E
D E
A B C E
B C E
A C E
C E
A B E
B E
A E
E

Arranged in alphabetical order, the list becomes:

A B C D E
A B C E
A B D E
A B E
A C D E
A C E
A D E
A E
B C D E
B C E
B D E
B E
C D E
C E
D E
E

It is interesting to observe that Venn diagrams, which are often used to demonstrate the principle of concept coordination, can be used to illustrate the principle of the SLIC index. In Fig. 2, three circles are used to represent



```
                                          MEMORANDA INDEX                           PAGE 93
KNITTING             WARPING              WRAPPING
     1345
KNITTING             WARPING              YARNS
      101 763 945 1181 1452 1565
KNITTING             WARPING              30 DENIER
      901
KNITTING             WEAVING
     1370 1601
KNITTING             WEAVING              YARNS
      763
KNITTING             WEFT
     1391 1672
KNITTING             WELDING
        7
KNITTING             WELTS
     1398
KNITTING             WORSTED
     1272
KNITTING             WRAPPING
     1345
KNITTING             YARNS
       69 101 122 164 763 772 885 886 945 979 1113 1151 1173 1181 1268 1379 1396 1411 1452 1565 1620 1664
KNITTING             YARNS                 6 DENIER
     1516
KNITTING             YARNS                15 DENIER
      727 1616 1620
KNITTING              6 DENIER
     1516
KNITTING             15 DENIER
      727 956 959 1371 1374 1616 1620 1631
KNITTING             30 DENIER
      901
KNOTS                POLYMERS             STRENGTH             TWINES               VISCOSITY
     1454
KNOTS                POLYMERS             STRENGTH             VISCOSITY
     1454
KNOTS                POLYMERS             TWINES               VISCOSITY
     1454
KNOTS                POLYMERS             VISCOSITY
     1454
KNOTS                RATIOS               ROPES                TENACITY
     1387
KNOTS                RATIOS               STRENGTH             TWINES
     1454
KNOTS                RATIOS               TENACITY
     1387
KNOTS                RATIOS               TWINES
     1454
KNOTS                ROPES                TENACITY
     1387
KNOTS                SLIPPING             WETTING
     1438
KNOTS                STRENGTH             TESTING              TWINES
     1438
KNOTS                STRENGTH             TWINES
     1438 1454
KNOTS                STRENGTH             TWINES               VISCOSITY
     1454
KNOTS                STRENGTH             VISCOSITY
     1454
KNOTS                TENACITY
     1387
KNOTS                TESTING              TWINES
     1438
KNOTS                TWINES
     1438 1454
KNOTS                TWINES               VISCOSITY
     1454
KNOTS                VISCOSITY
     1454
KNOTS                WETTING
     1438
KRALASTIC            SLEEVES              SNATCHING            TAKE OFF             TENSION
     1172
KRALASTIC            SLEEVES              SNATCHING            TENSION
     1172
KRALASTIC            SLEEVES              TAKE OFF             TENSION
      119 1172
KRALASTIC            SLEEVES              TENSION
      119 1172
KRALASTIC            SNATCHING            TAKE OFF             TENSION
     1172
```

Fig. 1. Sample of finished index.

three terms: A, B, and C. The several areas created by the circles are numbered 1–7. It can be seen that the total number of areas is equal to the total number of combinations $2^n - 1$, i.e., $2^3 - 1 = 7$. The selected combinations comprising the entries in a SLIC index using 3-term sets are represented by all those areas which fall within the last-named circle, i.e., Circle C. These areas (4, 5, 6, and 7) are four in number: $2^{(n-1)} = 2^{(3-1)} = 4$. This principle can be used for any number of terms and a 4-term diagram would show the combinations listed in the body of this paper. It is, unfortunately, impossible to use simple diagrams for more than 4 terms.

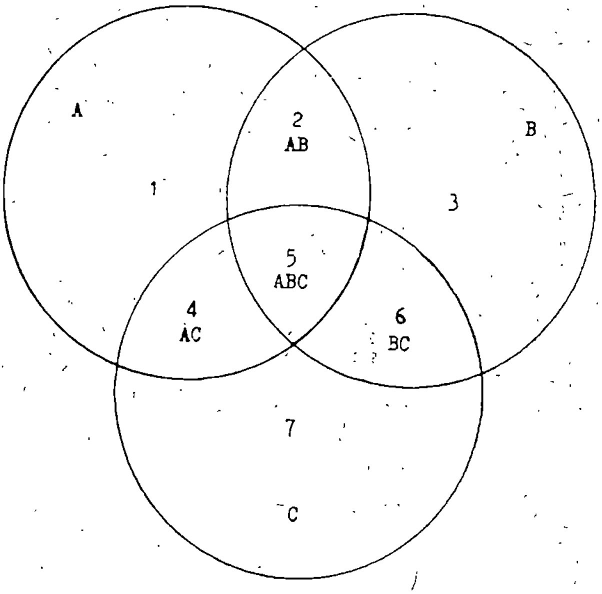


FIG. 2. Principle of the SLIC index.

Letters to the Editor

Dear Sir:

In spite of criticism and the pressures of suggested alternatives, scientific and technical papers, issued in serial format, are likely to remain the optimum technique for publication of most of the new research and development results, state-of-the-art and review presentations (1). Data suggests further, that as the number of such papers grows, repackaging into narrower segments through the splitting of old journals and the founding of new ones creates surges in circulation which indicate the appropriateness of this technique in enhancing the viability of the scientific serial.

Critics of the scientific serial condemn its distribution of too many papers of too little interest to any one reader, and propose complete disintegration of the mechanism through the establishment of separate distribution systems relying on demand orders from scientists who would select suitable papers from widely circulated title lists or abstract journals (2). This is the technique used to a considerable extent in the distribution of scientific and technical reports, and has not proved wholly acceptable. For one thing, it requires the reader to switch from the passive to the active mode in the dissemination system, and to operate this way continually.

Critics of the separates distribution proposals cite as advantages of the serial its process of preselection for the reader, its compensation in terms of prestige for editors in return for their considerable efforts, and its suitability to the presentation of advertising which allows publishers to tap this significant source of funds (3).

What is required is a system which would minimize the disabilities of both the separates and the serial mechanisms while retaining the principal advantages of both. Here, in brief, is a proposed system which may serve. In summary, it is a suggestion that serial publication be retained, but with varying groupings of papers within each series title, tailor-made to match smaller segments of readers than ever before through the use of selective dissemination of information (4) and automated typesetting and sorting techniques. In the proposed system, for any one journal:

1. Papers would continue to be written, refereed, and edited as at present.
2. Papers would be classified by subject to several levels of specificity using a limited number of terms (descriptor or microthesaurus concept).
3. The text of papers and identification numbers for each paper would be stored on paper tape. Identification numbers of papers and the subject terms describing each would be stored in computer memory.
4. Subscribers would be given identification numbers and would prepare profile descriptions, also to be stored in computer memory.
5. The computer would be programed to process its store at predetermined economic intervals. The output would be a list of subscribers' identification numbers, the identification numbers of the papers matching each subscriber's profile, and a tabulation of the number of copies of each paper required.
6. The tape containing the texts of articles, coupled with the print order for the number of copies of each paper required, would "drive" the typesetting and printing systems.
7. Papers would be automatically sorted into packets, each packet corresponding to the individual subscribers' profiles, as indicated by the printout in Step 5.
8. Advertising, news and editorial sections, and a title page and masthead would be "wrapped around" each packet; the entire packet would be covered, glued, and mailed.

The computer could be used for other control and service functions. If the number of papers for any one subscriber were too few, according to a predetermined economic issue size, the computer could select the appropriate number of additional papers from a more general level of specificity in the subject file. If too many papers were listed for an issue to a subscriber, the computer could provide a cutoff mark at the economic limit and provide a printed list of the titles of the remaining papers which could be sent to the subscriber who could either order full-text copies as separates or wait for his next journal issue when the full text would be scheduled to appear. Frequent occurrence of overselection or underselection in this part of the program would be an indication that the journal issue size should be changed for the next subscription period.

Each author could be provided with a list of the names or identification numbers of subscribers who had received his paper in order to be aware of his "audience."

Libraries would subscribe to serial sets of all papers to be filed under journal title by paper identification number. Library resistance to separates handling could be overcome by packaging papers into serially numbered groups. Bibliographical citation standards would require the use of the journal title and paper number. The paper number could be designed to indicate the year of publication or the volume number. Abstracting and indexing could take place as soon as a paper was selected for publication.

Selective dissemination would allow a greater number of different papers to be included in the journals than is now the case. Steps 6 through 8 would require technical developments in materials handling. There may, however, be other automated processes for serial publication, using alternative configurations of hardware, not here considered, which would give the same desirable effect.

References

1. Columbia University, Bureau of Applied Social Research. 1958. *The Flow of Information Among Scientists: Problems, Opportunities and Research Questions.* 175 pp. and append. New York.
2. Phelps, R., and Herlin, J. P. 1960. Alternatives to the scientific periodical. *UNESCO Bull. Libr.,* **14**(2): 61–75.
3. Thompson, G. P., and Baker, J. R. 1948. Proposed central publication of scientific papers. *Nature,* **161** (4098): 771–772.
4. Luhn, H. P. 1958. *Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment.* 19 pp. ASDD, IBM, Yorktown Heights, N. Y.

Russell Shank
Senior Lecturer
Columbia University
New York, N. Y.

Dear Sir:

The Golden Rule, as interpreted by publishers, seems to require them to avoid inflicting information on any seeker who might have to use it. By extension, we must never allow anyone to find information for which he may have to go through the motions of searching.

The Yellow Corallary holds that an author is ill-advised to make his sources findable, since someone may check up on him.

A few hateful scabs may actually wish to avoid bibliographic suicide, and run the risk of having orders come in from the book-hungry public. This may interrupt the publisher's poetry writing, and it may change his income-tax bracket, but some adventurous or acquisitive souls may be willing to face it.

The accompanying checklist can be used by either group, as a guide to bibliographic suicide or as a chart showing the rocks and shoals through which a newly-launched organization or publication may have to navigate.

This effort is not so much intended to discourage bibliographic suicide as to make it more efficient, where it is deliberate, or to permit people to avoid it, if they are so inclined.

Bibliographic Suicide-Guide: For Publishers and Authors

GENERAL RULES:

1. *Get people HEGGS!* HEGGS is "Eggs-aspirated."

2. *Give people ASININE!* ASININE means "Alphabet Soup of Initials with No Interpretation for the Non-Expert."

3. *Start your title with deadwood words. Journal of the . . .* offers so many more filing complications than, say, *Steel Research Journal.*

4. *Make sure that the initials of the name of your group or organization are embarrassing enough so that you will have to change to something and complicate life nicely for everyone.* The Society for the History of Operative Technology is ideally worded from this point of view. After a few years you can change it and get off all those pesky mailing lists.

5. *Start your title with initials.* The *ASTM Bulletin* is a classic of unfindability, especially after the clever change from American Society for Testing Materials became the American Society for Testing *and* Materials, which put everything just subtly out of place. The *IRE Transactions on . . .* and the *IEEE Transactions on . . .* are beautifully unfindable, since no one can guess whether to find them under the letters as one word or as though each letter were a separate word.

6. *Start your title with an easily misspelled word. Haematology Abstracts* would do nicely, since almost all Americans would look for *hematology* and the Britishers would look for the other. Right there you have managed to sidetrack half of the searchers, or prospective subscribers.

7. *Use confusing opening words. Committee* and *institute* are good, since no one can guess whether the abbreviation should be reconstituted as Committee or Commission, or Institute or Institution.

8. *Name the issuing body in the title of your journal. Journal of the Oobleck Research Society* is even better than *Oobleck Research Society Journal,* because starting with *Journal* throws off the people who might look for it under J rather than O, but it throws them further off.

9. Above all: *Never give a searcher an even break.* McGurk's Law holds that "Whatever would maximally foul things up, is maximally likely to happen." Make *sure* that your publication is *absolutely* unfindable or you will have to stop long enough to reconfuse the book-starved public which gets through to you.

ROBERT L. BIRCH
Science Index Group
3108 Dashiell Road
Falls Church, Virginia 22042

Erratum

American Documentation, Vol. 16, No. 4, October 1965, page 340: The letter concerning Chemical Abstracts Services and Chemical Titles is erroneously attributed to G. Salton. The writer is Dr. F. A. Tate, Chemical Abstracts Service, Columbus, Ohio.

Book Reviews

1/66–1R **A Directory of Information Resources in the United States: Physical Sciences, Biological Sciences, Engineering.** 1965. (Issued by the Library of Congress, National Referral Center for Science and Technology). U. S. Government Printing Office, Washington, D. C. 352 pp.

It is difficult to appraise a compilation of this kind, particularly when personal biases and prejudices are likely to color the reviewer's attitudes. But an objective evaluation of this book, after four months of actual use and testing (and some correspondence about particular collections), leaves one with a sense of the volume's potential helpfulness, especially when used with other similar tools — which is only to say that there is considerable difficulty in maintaining a complete record of organizations in the ever-changing fields of the biological and physical sciences and engineering.

Alphabetically arranged by name of organization, this directory gives addresses and telephone numbers, the latter a particularly useful feature, as anyone will know who has tried to locate a project of a subdivision of a subdivision of a university department with the aid of a long distance operator and the assistance of the university's telephone personnel. For each entry there is a descriptive note, of varying detail and length, which tells the scope or interests of the organization, but nothing to suggest its size; also, the quality of these descriptions varies considerably. Often there is a statement telling that book, journal, document, or report collections are maintained, although these are seldom qualitative evaluations and almost never indicate quantity (that is, size of the collection). Misleading or ambiguous statements — the mention of a collection — suggest that these materials are in some ordered arrangement and accessible. Interlibrary loan or photocopying policies, or rules about the availability of materials for the use of outsiders, are generally indicated. The imperfection of all these descriptions may lie, of course, in the character of the questionnaire, or in editorial revision of the returns.

The book's index is very useful because of the specific subject headings that have been used, but because there are too few headings and indexing has not been full enough, many possibilities are missed. For example, Ultra-Violet Products, Inc., of San Gabriel, California, says that "Books, journals, and reports are collected on ultraviolet and black light and fluorescent materials and related equipment. Chemicals, inks, and additives are manufactured for use with ultraviolet lamps." An entry like this might make an indexer's fingers itch and the p-slips fly, considering that the book is intended to be "A Directory of Information Resources," but the only entry here is under ULTRAVIOLET and other possible terms either do not include this organization, or do not appear at all in the index.

My chief complaint about this book and several others like it is the a priori method of compilation. As the "Foreword" says:

> This directory has been drawn from the central register of information resources being built up on a continuing basis by the National Referral Center. Because that register is still evolving and growing rapidly at the time of publication, the directory itself must be considered as a preliminary and exploratory effort rather than as a properly comprehensive guide, or even a properly selective one. Many resources of known significance have been omitted, for lack of data or other reasons; some resources of uncertain value have been included because of that very uncertainty; descriptions of services and functions are frequently less clear than could be desired, for lack of definitive terminology. These and other constraints must await future correction.

Apologetic explanations of this kind are acceptable but are of little help to the researcher who is in need (and how frequently we are told that scientists must have their information immediately). Of the private associations or corporation "libraries" listed here, I have personally investigated several that I had not known about before, and the "information resources" I saw were certainly not libraries in any sense of the word, nor were some of them any other kind of fount of information! In most cases there were no regular book collections — indeed, few had more than a half-dozen shelves of common textbooks — and there was no semblance of a catalog of any internal resources such as books, vertical file materials, etc. There were, in some cases, technical information reports of various series, but these were seldom unusual, unique, or really selective by content (though often only the "latest" reports are kept).

As for the special information that *is* available, it is either considered restricted information (by reason of government contract or "only for our own staff"), or it is in the head of one or two scientists or executives whose time or willingness to share their knowledge is limited by various factors. Further, one infrequently finds a professional librarian, documentalist, information scientist, archivist, or any other kind of specialist in charge of the arrangement, maintenance, or care, of these information resources. There is, rather, considerable chaos and very little sophisticated organization of special files or book collections in many of these places. Even when specialized series of documents are received and have a built-in, easy-filing classification and notation system, the filing is put off and the materials remain scattered throughout the organization.

This book, then, can be looked upon as a preliminary directory of some kind. It is certainly not a guide to any kind of library resources — and it is not meant to be, really — but one wonders whether it can actually be of much service to the people it is made for, even if it should appear in a more nearly perfected form in later editions. One thing is certain, it will help all kinds of people increase their mailing lists for one reason or another.

Of course one feature of the index is likely to be helpful, even if only by chance: instead of the usual interminable citations to page numbers (usual in books of this kind), the names of organizations are given; in this way, by scanning a list under a subject one might possibly find the name of an organization where a former employee, a college chum, or a conference drinking companion works. This may be especially helpful because so much specialized, confidential, and even restricted information is passed on through the buddy-buddy association of "contacts." In any case, users should remember again that this is not a directory of libraries but of organizations, meaning that inquiries may not always be received or answered with the service-oriented amiability we are used to finding among librarians.

As the compiler of another kind of guide to library resources, the reviewer recognizes that he and the National Referral Center people have many problems in common. It is his ardent hope that the NRC will, by constant effort and careful study, come up with some easy way to describe resources accurately and to appraise the willingness and ability of organization personnel to share their stored information.

LEE ASH
Compiler, Subject Collections:
A Guide to Special Book
Collections and Subject
Emphases

1/66–2R **Author-Title Catalogue. Subject Catalogue.** April 1965. Toronto University Library, Ontario New Universities Library Project. 2 vols.

Catalog of Books Plus a Complete Catalog of Reserve Books. 1965. Washington University School of Medicine Library, St. Louis, Mo. 267 pp.

In recent years, advances in technology have begun to make book catalogs in libraries a reasonable alternative again for the first time since the advent of Library of Congress printed cards. As a result, a number of libraries, new and old, have begun to experiment with catalogs in this format.

The Ontario New Universities Library Project (ONULP) and the Washington University School of Medicine Library catalogs are two recent examples of this trend. Both are compiled on the computer and use IBM printouts, and both are divided, but there the resemblance ends.

The ONULP project is intended to provide five new universities with identical basic collections of about 40,000 volumes by 1967. The book catalog represents only these collections and excludes other materials which the libraries will acquire. This catalog is being published monthly, with quarterly and semiannual cumulations and annual total cumulations. It uses a specially-developed 120-character upper and lower case print chain with diacritical marks, and is divided into two parts: author-title and subject. The format is attractive and the 42% reduction ratio produces quite legible copy.

The catalog is compactly arranged, without much wasted space in the author-title catalog. Cataloging information is given in full in the main entry but is abbreviated elsewhere. However, there are three instances where some wasted space occurs which might bulk large when the catalog reaches its projected 1967 size. Filing titles are used in the Shakespeare entries and appear in the printout. Since we are dealing with a union catalog, the line devoted to location designations for each entry appears reasonable at first, but the preface explicitly states that only the identical basic collections are included, and every entry turns out to have the location symbol for all five libraries.

In the subject catalog there are almost as many cross references as there are subject headings with entries under them. Many, if not most, of these could easily have been omitted: on the first page the entry. *Abailard, Pierre, 1097–1142,* is surrounded by cross references from five other forms of the name. There are no intervening entries. This case, while extreme, is not atypical.

The arrangement is entirely by computer, using this order of sorts: blank, period, dash, comma, A through Z, 0 through 9. Thus far, provision has been made for disregarding initial articles in title entries, and qualifiers such as *ed.* in name entries.

The clerical portion of the work is very well performed: there are no peculiarities of spacing to lose entries or make them fall into two files. However, the mixture of open and closed author's dates, and the evident failure of the computer to disregard *ed.* when it follows a person's dates, might well cause problems when the catalog has reached its projected size.

The practice of filing on punctuation was probably followed because it does solve some difficulties, particularly in subject entries. However, it must require extra care in cataloging and punching and produces some peculiarities. For instance, in all subject headings which are subdivided, the punctuation mark is both preceded and followed by a space:

Meteorology , Agricultural
Athens . Theater of Dionysus
Art - Africa , South

The ONULP catalog certainly did not create the problem of subject headings, but it demonstrates them beautifully. The entries under *Gt Brit* - History, shown below, are a mixture of form and dated and undated period subdivisions led in a strict sort order bearing no relation to the separation of form and period divisions, with chronological arrangement of the latter, which is usually followed.

Gt Brit - History
- Addresses, essays, lectures
- John, 1199–1216
- Medieval period, 1066–1485
- Stephen, 1135–1154
- To 1485
- Tudors, 1485–1603
- Tudors, 1485–1603—Sources
- Victoria, 1937–1901—Addresses, essays, lectures
- 16th century
- 1689–1714
- 1714–1837
- 1760–1789
- 18th century
- 20th century

, Economic
, Local
, Military
, Political

These headings now occupy about 2½ columns in a subject catalog of just under 2,000 entries. When the libraries have reached their projected 1967 size of 40,000 volumes, arrangement of subject headings will be a major problem if their form is not changed before then.

As in any human effort, the ONULP catalog has its share of typographical errors (but not very many) and spots where the gods nodded and forgot that computers are logical beasts. But to detail these would be hairsplitting, to no point. The catalog will fulfill most adequately the purpose for which it is intended, and the introduction makes evident the awareness that it is a first, experimental effort, and that changes are planned.

The Catalog of Books Plus a Complete Catalog of Reserve Books of the Washington University School of Medicine Library is not that at all. It is a catalog of accessions in the Library from January 1 through September 1, 1965, consisting of 1,050 items. The most unfortunate thing about it is the statement in the preface that the Library is not planning to continue the printed book catalog. While it is possible that the reasons given for this decision might be valid (production costs and lack of demand for the previously-produced, semiannual, cumulated serials holdings lists), this catalog certainly will not provide a fair demonstration because it covers only a very small percentage of the available material.

The catalog is quite valuable as an experimental demonstration. It is part of a computer-based system set up at the Library to make one input manipulable for numerous outputs, from acquisitions to cataloging. The results of the experimental work have been well reported in the literature, including costs, so need not be reviewed here.

The catalog is divided into four sections: Author and added entry, Title and series title, Subject, and an author listing of the Reserve collection. (The last of these, separately published, might have provided some idea of the real demand for a book catalog.)

The prefatory material discusses several problems that arose during the making of the catalog, and that are present in this version. Among these are an oversight, the result of which was that in the author and subject catalogs many of the entries by the same author are not subarranged by title at all, and the remainder are alphabetized only by the first letter of the title. In most cases this is a minor problem, but in some of the U. S. entries, with up to 19 items under a single author, not subarranged at all, the story is different.

Another peculiarity is the substitution of cross references for added entries in the author file. Usually this is not a major problem, but when reference is made to an author heading under which there are several entries the entire group must be scanned to find the relevant title.

An article describing the projected catalog in the *Bulletin* of the Medical Library Association states that the subject headings were coded for arranging purposes. However, it is difficult to imagine what coding procedure could have resulted in an entry under DAIGNOSIS, filed between CYTOXAN and DARWIN, plus two entries under DIAGNOSIS; between DIABETES MELLITUS, JUVENILE and DIARRHEA.

One serious mistake was use of the direct printout without reduction in size, so that the catalog, which totals about 4,000 entries, is 267 pages long, 8½" × 11". The subject catalog, with 113 pages for only 1,235 entries, is the worst offender.

This catalog is admittedly experimental. The conclusions reached as a result of its production, and stated in the prefatory material, could well have made possible the subsequent production of an improved, more complete catalog that would have served a real purpose for users of the Library, thus providing a fairer test of the system. It may be hoped that the Library will reconsider its position and attempt a more complete catalog in the future, especially since all the data will be on tape anyway.

Jessica L. Harris
Rothines Associates

1/66-3R **The Education of Science Information Personnel — 1964.** 1965. A. J. Goldwyn and A. M. Rees, Eds. Western Reserve University, Cleveland. 115 pp.

This book presents the proceedings of a two-day conference held in July 1964. It consisted of five sessions, the crucial one being the first:

1. Summary statements by 17 colleges and universities describing their programs, approaches, and attitudes concerning the education of science information specialists and information scientists. Of the 17, 14 are library schools and 3 are not. Of the 17, 9 have no active program beyond a single course in "Documentation" or "Information Retrieval," 4 have emphasized some kind of documentation program designed for information specialists, and 4 have a defined program in information science as a more or less theoretical discipline.
2. Three papers on the general topics of manpower and research. The first two papers, on manpower, were presented by Robert Kohn and William Hitt of Battelle, and emphasized their study of the needs for and use of manpower in engineering and the natural sciences. Mr. Kohn discussed the intent and approach; Dr. Hitt presented the methodology and 10 fundamental questions to be answered (What is the field? The job function? Routes of entry? Characteristics of personnel? Skill shortages? Educational needs? Etc.).
3. A presentation by Stafford L. Warren of his proposal for a Library of Science System and Network, which would use Medlars as a starting point.
4. A series of workshop reports on students, faculty, curricula, and academic organization. The magnitude of the problems in these four areas is so great that it is a pity the workshops did not produce more than the limited results reported in this section. However, these results reflect the limitations of conferences more than the interest and capabilities of the participants. Such workshops may represent useful educational experiences for the participants; reports of them rarely are useful, and these are no exception.
5. A brief summary by A. J. Goldwyn.

The results of this conference, in comparison with its predecessors at Georgia Tech., reflect the extent to which curricula have been formalized throughout the United States. The last few years have seen progress in at least three areas of educational programs: instruction in library automation, education of information specialists, and development of research programs in information science. The reports in this book show the magnitude of these developments.

Robert Hayes
School of Library Service
University of California at
Los Angeles

1/66-4R **Technical Dictionary of Librarianship, English-Spanish.** 1964. Beatriz Massa de Gil, Ray Trautman, and Peter Goy. Editorial F. Trillas, S. A., Mexico. 387 pp.

In this work the authors set out to design a dictionary "for librarians, students, editors, publishers, booksellers, archivists, and all others who work in the communication arts or are interested in the technology of librarianship." Since there is a scarcity of reference books of this nature, there is no doubt that the Spanish-English and English-Spanish vocabulary of over 3,000 terms will be helpful to those for whom the dictionary is intended. Nonetheless, the prospective user should not become overly enthusiastic and expect to find much of the technical terminology which has appeared in the last few years. It is the traditional aspects of librarianship which are emphasized, perhaps at times to the point of redundancy. Words such as BIBLE, PARAGRAPH, PENCIL, SONG, and SCIENCE, that can readily be located in other standard bilingual dictionaries could have been omitted to make way for RETRIEVAL, DESCRIPTORS, INDICATORS, and the many other terms and expressions that have developed with data processing.

The arrangement of the material is laudable. It is divided into two parts: Spanish-English and English-Spanish. In each section the order is alphabetical word by word. The lexicographers have not merely translated each term from one language to the other, but have gone a step further by defining the vocables included. However, since the historical and/or geographical backgrounds of the words are not included, the reader should realize that there can be variation in the meaning from one country to another. This is especially true in Spanish-speaking countries where regionalisms can persist as a result of the limited interchange of professional literature. In Mexico, for example, *encabezamiento de materia* is the commonly accepted designation for "subject heading," yet in Peru the term used is *epigrafe de materia.*

All in all, this technical dictionary is a definite contribution to the field of library science. It should do much toward the systematization of its lexicography. It is to be hoped that more books of this type will be forthcoming in the near future.

Arnulfo D. Trejo
School of Library Service
University of California at
Los Angeles

Documentation Abstracts

. . . . . is a joint publication under the auspices of the American Documentation Institute and the Chemical Literature Division of the American Chemical Society.

. . . . . represents combined coverage of the former Literature Notes section of American Documentation, the ACS Division of Chemical Literature Annotated Bibliography, and the former coverage of Documentation Digest.

. . . . . will issue quarterly — February, May, August, and November of 1966. Each issue will contain corporate and author indexes; subject indexes will be available on a schedule to be determined.

Subscriptions are sold on a calendar year basis — $8.00 per year.* Return the coupon below. Payment with your order is requested.

* Members of the American Documentation Institute will receive the first year's subscription free.

DOCUMENTATION ABSTRACTS—Please enter my subscription for one year commencing with the February 1966 issue. At $8.00 per year, payment is enclosed □ bill me □

Name______________________________ Title______________________

□ Business
Address □ Home ____

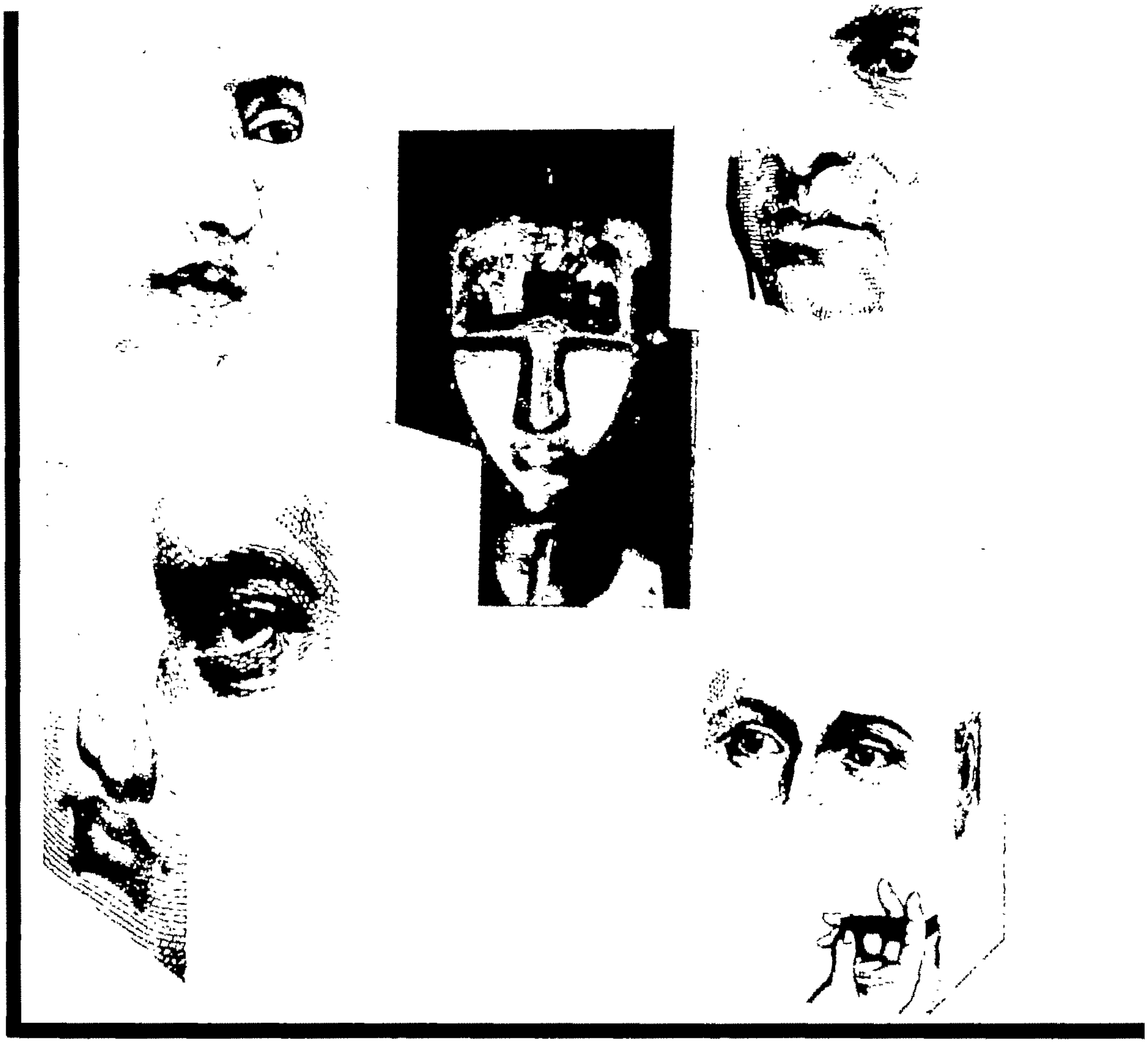
City _________________________ State________________ Zip______

Your Firm Name ____

merican Documentation

APRIL
1966
Vol. 17, No. 2

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE



AMERICAN DOCUMENTATION

INSTRUCTIONS TO AUTHORS

American Documentation is a publication of the American Documentation Institute. It is a scholarly journal in the various fields in documentation and serves as a forum for discussion and experimentation. Papers already published or in press elsewhere are not acceptable. For each proposed contribution, one original and two copies (in English only) should be mailed to Mr. Arthur W. Elias, Editor, *American Documentation,* Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106. The manuscript should be mailed *flat* in a suitable-sized envelope. Graphic materials should be submitted with suitable cardboard backing.

Types of Manuscripts: Three types of contributions are considered for publication: full-length articles, brief communications of 1,000 words or less, and letters to the editor. Letters and brief communications can generally be published sooner than full-length manuscripts. Books, monographs, and reports are accepted for critical review. Two copies should be addressed to the Review Editor, Dr. T. Hines, 54 North Drive, East Brunswick, New Jersey.

Processing: Acknowledgment will be made of receipt of all manuscripts. *American Documentation* employs a reviewing procedure in which all mansucripts are sent to two referees for comment. When both referees have replied, copies of their comments are sent to authors with the Editor's decision as to acceptability. The refereeing procedure requires about 30 days. Authors receive galley proofs with a five-day allowance for corrections. Standard proofreading marks should be employed. Reprint order forms are forwarded with galleys.

Format: All contributions should be typewritten on white bond paper on one side only, leaving about 1.25 inches (or 3 cm) of space around all margins of standard, letter-size (8.5 × 11 inch) paper. Double spacing must be used throughout, including the title page, tables, legends, and references. The first page of the manuscript should carry both the first and last names of all authors, the institutions or organizations with which the authors are affiliated, and notation as to which author should receive the galleys for proofreading. All succeeding pages should carry the last name of the first author in the upper right-hand corner (0.5 inch from the top) and the number of the page.

Style: In general, style should follow the forms given in the Style Manual for Biological Journals (SMBJ), published for the Conference of Biological Editors by the American Institute of Biological Sciences (1964).

Title: The title should be as brief, specific, and descriptive as possible. Vague and unrevealing titles may delay publication.

Abstract: An informative abstract of 200 words or less must be included, typed with double spacing on a separate sheet. This abstract should present the scope of the work, methods, results, and conclusions.

Acknowledgments: Financial support may be listed as a footnote to the title. Credit for materials and technical assistance or advice may be cited in a section headed "Acknowledgments," which should appear at the end of the text. General use of footnotes in the text should be avoided.

Graphic materials: *American Documentation* requires finished artwork. Follow the style in current issues for layout and type faces in tables and figures. A table or figure should be constructed so as to be completely intelligible without further reference to the text. Lengthy tabulations of essentially similar data should be avoided.

Figures should be lettered in black India ink. Charts drawn in India ink should be so executed throughout, with no typewritten material included. Letters and numbers appearing in figures should be distinct and large enough so that no character will be less than 2 mm high after reduction. A line 0.4 mm wide reproduces satisfactorily when reduced by one-half. Graphs, charts, and photographs should be given consecutive figure numbers as they will appear in the text; however, figure numbers and legends should not appear as part of the figure, but should be typed double spaced on a separate sheet of paper. Each figure should be marked *lightly* on the back with the figure number, author's name, complete address, and shortened title of the paper.

For figures, the originals with two clearly legible reproductions (to be sent to referees) should accompany the manuscript. In the case of photographs, three glossy prints are required, preferably 8 × 10 inches.

Organization: In general, papers should state the background and purpose of the study, followed by details of methods, materials, procedures, and equipment. Findings, discussion, and conclusions should appear in that order. Appendixes may be employed where appropriate for extensive lists, statistics, and other supporting data.

Bibliography: Accuracy and adequacy of the references are the responsibility of the author. Therefore, literature cited should be checked carefully with the original publications. References to personal letters, abstracts of verbal reports, and other unedited material may be included. If an as-yet-unpublished paper would be helpful in the evaluation of a manuscript, it is advisable to make a copy of it available to the Editor. When a manuscript is one of a series of papers, the preceding member of the series should be included in literature cited.

Citation Format:

Order: Literature cited should be sequentially numbered as cited.

Authors: Give all authors with arrangement as follows:
Elias, A. W., B. H. Weil, and I. D. Welt

Titles: Give full titles of articles in English, indicating language of original as: (In Ger.)

Journals: Journal titles should be given in full.

Monograph and Serial Data: Should be presented in order as follows: Volume, issue number, pagination, and year. The issue number should be given in parentheses if journal pagination is not continuous from issue to issue. Pagination should be inclusive. Year of publication should be given in parentheses. An example is given below:

Bishop, D., A. L. Milner, and F. W. Roper, Publication Patterns of Scientific Serials, American Documentation, 16 (No. 2): 113–21 (1965).

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), three copies in the contributing membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts, Science Citation Index, Chemical Abstracts,* and *Documentation Abstracts.*

American Documentation is entered for second class mailing at Baltimore, Maryland.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 17, No. 2 APRIL 1966

Published quarterly by AMERICAN DOCUMENTATION INSTITUTE and printed by Port City Press, Inc., 1323 Greenwood Road, Baltimore, Maryland 21208. Second class postage paid at Baltimore, Maryland.

Pop Science

DANIEL I. COOPER

International Science and Technology

EDITOR'S NOTE: American Documentation *is privileged to publish these remarks. They have been made available as the text of a presentation delivered at the Gordon Conference on Scientific Information in July 1965.*

Of course I chose my title by analogy with Pop Art.

Paintings of endless rows of Campbell's Tomato Soup, three-dimensional *trompe l'œil* Brillo boxes — what strange pieces these are. How unattractive, really. I suspect most people don't like Pop Art. And of course the remarkable thing is that the artist paints them to *be* objectionable. He waits to so engulf us with the banality of our society, with its misuse of art, that we will *do* something to change the world or at least to change ourselves.

The Pop Artist — the good one — is trying to rub our nerve ends raw so that even as lovely a piece of commercial photography as a well endowed 36–26–36 in a four poster will still rouse some objection in us.

She's lovely, isn't she?

Makes you feel warm inside.

But she's not there for the purpose you have in mind.

She's there to sell cosmetics. Silicone-based cosmetics at that, — a product of the silicone chemists at Union Carbide. So there is one connection — however tenuous — between pop art and pop science.

But I don't want to pursue *that* connection. Indeed this entire prologue is designed only to restore ambiguity and multivalued criteria to what I'm sure has been an orderly week with matters settled by the strict rule of the scientific method.

My text for this evening's sermon — if sermon it be — comes from the French mathematician, Poincaré.

On fait la science avec des faits comme un fait une maison avec des pierres; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.

In English: One builds science with facts as one builds a house with stones, but a pile of facts is no more a science than a pile of stones is a house.

I wonder if it is so. Oh, it is not without truth — surely there *is* an elaborate and careful structure of the facts of science tied together with the theories of which a mathematician like Poincaré could be so proud.

But that structure is in good measure the work of, teachers, of the writers of review papers, of the rapporteurs at our ever more numerous conferences. The building of science is a more chaotic process. Long before one has the mansion of science one has the tas de faits . . . . the pile of facts.

To press my luck on the analogy: the building of science, so elegant, so overwhelming in its final form, passes through a stage when the workmen's materials are strewn about and the workmen themselves — dirty, sweaty, clad in over-alls — lounge in what will someday be the great court.

What in the world has this to do with *Pop?* Well what is more of the people — which is what *popular* means — than the workman. I see him in New York, muscles rippling, chewing on his hero sandwich, calling "Hey, goodlooking!" or "Bella! Bella!" to the lovely young things who pass below his noonday perch while we poor professionals can only stare in hungry silence as we trot off to our martinis. It's all pop. Or quack, as Saul Bellow has it in *Herzog*.

Well now, why divert you — near the end of a week's serious discourse on scientific information problems — with images of pretty girls and all that conjures up. Perhaps it's just because you have been serious — and abstract — that I feel the need to divert you. I want to return your attention to the individual acts of generating a new idea that sum up into the body of scientific information that has been your concern.

You see, the tradition has grown of making the research paper a smoothed-over, tight exposition of the end point to which a man's researches brought him. Agreed: As part of the discussion of where things stand, the paper may contain some allusion to the yet unresolved problems of the field. But no sense of the research itself, of the gnawing uneasiness, the discomfort the researcher experiences at having some part of this subject that he loves unclear. That's what drives him to seek clarification, to stay in the lab till all hours, to ignore his wife. Because the pleasure that comes *with* clarification surpasses all other pleasures.

Robert Wilson of Cornell, in an interview in our magazine, likened the whole experience to throwing up. This awful queasiness, this rumbling around inside, this subconscious knowledge that *something* is going to happen —

that the queasiness can't be maintained. This *body* knowledge: And then the relief — the resolution, the orgiastic moment of relief, the *pleasure* of having tension disappear — like finishing a speech, or finishing the preparation of it.

To do research is to be in a special state of grace. I've written of the special look I've observed on the faces of physicists at meetings of the American Physical Society. These are bearers of a special sort of knowledge, these are — in the main — happy men. They are priests of a true religion.

But such is the nature of many of these men, or such is their training, or such is the tradition of the scientific paper, that few of them *talk* about their state of grace. And fewer yet write about it.

I say we're the poorer for it. We're the poorer because writing is a means for sharing experience. Insofar as we share the smoothed-over, prettied-up, rationalized *maison de science* rather than the *tas de pierres* — to that extent we are all impoverished. I maintain that our archival journals suffer for not recording more of the raw experience. That's why conferences have become so popular even though oral communication is inherently less inefficient than written: At conferences such as this there is the man-to-man opportunity to confess error, to recount the blind alleys, to explore fresh paths together, to discover that we are not alone in our stupidities.

Now I'm not proposing that the archival journals become something akin to a poetry reading in a fifth-rate Greenwich Village Coffee Shop — all passion and no content. I don't mean for the *Journal of the American Chemical Society* to become *Chemical Confessions* or *Chemistry Confidential*. But I do ask for some leavening of some of the standard research papers with some of the *actualities* of the experiment, not the prettied up "Results" with their echo of our schoolroom experiments with their preordained outcomes. Not *Chemical Confessions* but *Chemical Candor*. Mind PLUS Emotion.

It's interesting that this sort of thing *does* take place with increasing though insufficient frequency in our review and interdisciplinary journals. Here is *Science* magazine carrying a debate about Superconductivity between P. W. Anderson and Bernd Matthias, both of Bell Labs. You can get some sense of the quality of this piece from the opening paragraph of Anderson's remarks:

> The conditions under which this article is being written are unusual. With the other side of the coin being ably presented by my colleague, B. T. Matthias, I will not have to qualify my statements or judiciously distribute credits and concessions, but can flatly state my opinions, for what they are worth. I suspect that I will be proved wrong in some measure; I hope the fact of my stating these opinions will stimulate other physicists to try to prove me wrong.

And the article follows in that spirit.

Notice another remarkable result of setting aside the normal, *formal* structure of the research paper for this debate in print. The paper is more *personal;* in the heat of preparing for the debate Anderson drops the customary (and awful) impassive voice and says "I." "*I* suspect," "*I* hope," "*I* will not"; in fact he uses one first person pronoun or another 9 times in a three-sentence opening paragraph. Surely some sort of record for a paper in a scientific journal.

Why, it is even more than I find in a typical paragraph of what is my favorite reading whenever I return to Boston: the Confidential Chat page of the *Boston Globe*. On that page sweet old ladies carry on public correspondence under equally sweet code names. Here's one:

> Chat Editor — I never miss an issue of the Chat, believe me. I pass it on to my daughter and she passes it on to a neighbor. Quite often there are things cut out before they get it, but they do not mind because they are glad to get it. Sunday's pages were a real bonus. Wish it could happen more often. I will return to my old pen name because since I dropped it I have never once seen it used.
>
> I have six dogs and a cat (they get along fine) and a very large vegetable garden besides sewing, knitting and canning. Life is beautiful, my days are full, never time for everything. Add to that the Boston Globe, what more do I need?
>
> —Hilltop

Now why do I bring that in? Listen to the lady:

> I have six dogs and a cat (they get along fine) and a very large vegetable garden besides sewing, knitting and canning.

She's proud of what she's doing — it's what she *is*. It provides — to use that horrible word — IDENTITY.

Susan Langer in the introduction to her *Philosophy in a New Key*, a fairly steep book about art and symbology, reminds us that all of art has much in common with the instincts that cause us to show off our mud pies when we are young.

I hope you won't forget as you edit scientific papers, then abstract them, then put the titles into a key-word index, then study the statistics of such titles, then prepare the whole for instantaneous electronic retrieval, and then hold conferences — pleasant conferences — on the whole subject. . . . I hope in all this you won't forget that you are dealing with men's passions and hopes, with a sometimes desperate attempt to leave some sort of scratch on the face of anonymity before death overtakes us all.

What I'm saying is that a catalog of Picasso's paintings, while necessary, is still not Picasso.

What I'm saying is that every bit of Scientific Communication is at the same time a showing off of a mud pie: "I have six dogs and a cat (they get along fine) and a very large vegetable garden besides sewing, knitting and canning."

What *if* the mud pie slumps!

What *if* the cat gets chased once in a while.

What *if* the paper ain't so vital.

It's *my* mud pie, *my* pets, *my* paper.

So my plea is for a recognition of the *pop* in science and for a search for ways to convey it.

We've found one way in *International Science and Technology:* a sort of interview in which good scientists and engineers talk in a very personal way about how they do science and what it means to them:

Bernd Matthias of Bell Labs in an interview entitled *The Gambler in the Laboratory:*

> You see, I like to gamble, and I do the same thing in physics. I look for new things. If you do this, there are three possibilities. Either you find what you are looking for, or you find something else, or you don't find anything. If you don't find anything, there is nothing you have to show for it. Oh sure, today people want to publish negative results, but it is always an anticlimax. I'm quite willing to gamble. If I find things, fine; if I don't, well, I've just lost.

Bob Wilson of Cornell, speaking of *The Pleasures of Physics,* telling how he very nearly won World War II singlehanded:

> It was late 1941 — right in the worst moments of the war. The time of Pearl Harbor. The Battle of Britain was just over, but things were still at a very low point. If one could make a bomb, that would be the salvation of the world, not the damnation of it.
>
> So with this idea and in that desperate situation, I thought, "My God! I just have to learn how to separate isotopes!" I thought long and hard — intensely for a number of days. My thoughts turned to the electrical methods. I can still remember vividly the clear cold air and the experience of walking through it and thinking "By God, it's going to come." And come it did. I was conscious that the idea was there within me before it finally revealed itself to me. That idea subsequently became known as the Isotron.
>
> I became extremely excited, and, as I walked along, my ego became all involved. I felt, I, a young man of about 25, would almost personally win the war. In a few months, if we worked hard, we at Princeton could test this thing, we could get a few grams of U-235, and then we would make a bomb and stop the war. And we could have. It *was* possible — had the neutron cross sections worked to be larger, as the British thought, it *could* have happened.

The mud pie just slumped — but what a mud pie . . . what a moment for any man to experience.

Here is F. C. Williams of Manchester, the inventor of the Williams Storage Tube, telling all of us off in an interview entitled *How to Invent.*

> I think it's a great mistake to learn too much, to be taught too much, to be too good at anything, because this tends to become important in itself. It's just no good knowing about these things, if you're not going to do anything about them. You might just as well study Shakespeare and know all about that, because you're not going to do anything with that either. There's a great danger you know, that scientific education will go that way. That it will become a virtue within itself to be able to do things that can already be done — whereas the true virtue is to be able to do things that have not already been done.
>
> I would suggest, if you're trying to make progress nowadays in the computer business, you can do one of two things: You can either work on your own and try and make some progress, or you can keep abreast of what other people are doing. But you damned certainly can't do both. . . .
>
> There's only one easy place to be in science and in engineering, and that's in the front. If you're there first, you have nothing to read. You've got all your time to think.

Around now some of you have a perfect right to protest.

I can almost hear the thought waves: It's perfectly all right for you as the publisher of a pop magazine to plead for the pop in science, but I'm a librarian — excuse me, a documentationalist. My motto is Orderliness. My decalogue comes not from Moses but from Dewey. My mud pie is to keep other people's mud pies in straight rows.

Well, I would respond to the good lady (I know librarians are not always women, it's just more pleasant to think of them that way) by first pointing out my complete respect for her craft. Having demonstrated this evening that I can barely keep script and slides and references together, I trust I don't need to elaborate on how highly I regard someone who can keep all of the world's knowledge pigeonholed. But, you know, there is pop in the library, too. Sure the books are all in the stacks where they should be . . . (Or almost so. One of the delights of a library is coming on a good book on the Mexican War when you are searching in decimal classification 749.2 for a book that will tell you how to build that wall your wife wants. What a welcome diversion for a man whose mud pies — and walls — always slump!)

But by and large the stacks *are* orderly, so I shouldn't try and put you off that way. Rather I would make the point that there's a lot more going on in the stacks than you realize. People are *browsing.* The stacks are orderly, but our *minds* are not.

As long as you as a librarian keep your stacks and shelves open a lot of *pop* is taking place:

> What a nice book this is!
> What's that blue book about?
> I *like* blue bindings.
> What in hell do we get the *Journal of Oral Surgery* for?
> Etc . . . Etc . . . Etc.

Pop . . . Pop . . . Pop.

There's more. My librarian — back home in South Orange — usually has a suggestion for me. "You'll like this book, Mr. Cooper." "I didn't like *Herzog* but you might, Mr. Cooper." And because I can't say no to anyone, and because I know that the librarian is depressed because the town voted down a new library (three times now), I usually accept her suggestion and am glad for it.

What does this suggest for documentationalists? Well, maybe these things, and now I am serious — sort of:

(1) I worry when our journals, our abstract lists, our information-retrieval schemes are not in danger of becoming too efficient, too adjusted to what the customer is thought to need. I hope the narrowest will always keep some window on the world, some device for providing the unexpected, the pop, some provision for browsing.

(2) Specifically, I would hope that those of you who

write computer programs for information retrieval will provide for browsing in the reference lists thus prepared . . . maybe by having the machine read two numbers out of a random-number table and let the first signify the location and the second the title of a random (but good) reference that would stop your reader in his too avid pursuit of papers on the melting point of gallium. He may benefit so much more from knowing that new techniques exist for automating such measurements. Or he may discover that the galaxy he inhabits is bigger than had been thought . . . that sort of discovery puts the melting point of gallium into perspective, somehow.

(3) I don't see why abstract services and computer programs can't have opinions like my librarian. A nice little journal called *The American Behavioral Scientist* has a section devoted to abstracts of the literature. It's complete — I believe — but also evaluative. Abstracts that strike the editors as being of more than ordinary interest are surrounded by a box. And, thus highlighted, the reader is guided to the best of what otherwise would be a dull-looking list.

(4) Speaking of dull-looking, can't you folks get those computers to print out in more interesting type faces? Frankly one thing that keeps readers away from computerized abstracts in droves is the fact that they look like they were prepared *by* machines *for* machines. My whole point this evening is that we are *not* machines . . . that all this communication with which we are concerned is an attempt to transmute the Pop that had occurred in one man's mind into a Pop in the minds of his audience wherever they may be.

Just a word more on the meanings of Pop Science and why all this is so important.

First, as you all know, science is getting much bigger. Extrapolating ahead it would seem that all mankind will some day be scientists. If we are not to have a vulgarization, a degradation of science — a pop science now in the same sense that the pop artist is protesting the popular vulgarization of his métier — then we must communicate the spirit, the actualities of science as well as its facts.

Second, as science impinges on our lives more and more the temptation grows to misuse science for commercial gain. To use science to sell soda pop. I mean the irritating, false use of science in TV and newspaper advertising.

Finally there is science as Pop . . . Dad, Father, The Old Man. Let's face it — this is an *age* of science and technology. People like us develop the principal instrumentalities for our modern society, for this postcivilization, as Kenneth Boulding has called it.

It's terribly important that the technology thus developed not be mindless — that it be applied with due regard for the consequences. It's perhaps even more important that it not be *heartless*, devoid of passionate concern. That, really, is why I chose to carry you in this direction tonight.

The KWIC Index Concept: A Retrospective View

This paper defines and describes the KWIC (keyword in context) index concept, providing a history of the concept and of its literature. It discusses variations of the index, such as the Bell Telephone Index, KWOC indexes, and the WADEX.

The paper discusses improvements and variations to the KWIC index, such as manipulation of the index line, variations of the code, addition of classification information, combination of author index and title index, and improvements of the type face. It also discusses improvements to the preparation of the KWIC indexes, such as improvement of titles and use of a thesaurus, and discusses improvement of the use of the KWIC index. The paper discusses the usage of the KWIC index and comments on the future of KWIC indexes and of the KWIC concept.

MARGUERITE FISCHER *

American College of Physicians
Philadelphia, Pa.

• A Review of the Literature

The classic paper on the KWIC index is *Keyword-in-Context Index for Technical Literature (KWIC Index)* published by Hans Peter Luhn in 1959. This paper introduced the idea and the plan for a permutation index based on titles, and produced by machine.[1]

Earlier, at the International Conference on Scientific Information (1958), both Luhn and Ohlman [2] had distributed copies of machine prepared permuted indexes which each had developed independently.

The use of the KWIC index in the preparation of *Chemical Titles* was described in a comment published in *Law Library Journal* in 1960.[3]

Papers by Lester Douglas Turner and James Henry Kennedy, appearing in 1961, explained SAPIR (system of automatic processing and indexing of reports), which used the Keyword-in-Context index principle.[4] The same year an article by John H. Veyette, Jr., fixed the KWIC index in a pattern of information retrieval,[5] and an article by A. Resnick similarly bore upon an aspect of information retrieval as related to the KWIC system.[6]

By 1962 the KWIC index and variations of it had become widespread enough in use to occasion further explanation, evaluation, and criticism. For this period, the *General Information Manual* issued by IBM may be considered as an authoritative source of information concerning the KWIC index.[7] Instructions for production of KWIC indexes came from the Space Guidance Center, IBM, in a report by Charles H. Balz and Richard H. Stanwood.[8] At the same time, Frank V. Giallanza and James H. Kennedy wrote of the KWIT (Keyword-in-Title) Index used at the Lawrence Radiation Laboratory, discussing possible options for preparation of KWIC indexes of subject, author, report number, and field-of-interest.[9] At the University of Oklahoma, a research project was underway to use the KWIC program for retrieval of space law materials.[10] The American Diabetes Association was considering use of the KWIC index to provide one of a number of desired indexing depths.[11] Stanwood proposed the Merge system, a complete information system linking the techniques of Keyword-in-Context with SDI (Selective Dissemination of Information).[12] Library applications practiced at Bell Telephone Laboratories were reported by R. A. Kennedy.[13] Donald H. Kraft evaluated the efficiency of the KWIC principle on the basis of his study of legal document title entries.[14] William J. Kurmey undertook the comparison of keyword-in-context effectiveness with that of subject heading effectiveness.[15] In England, J. D. Black described the KWIC concept as offering advantages not possible with conventional indexing, and cited user reaction to KWIC as being favorable.[16] Mary Veilleux described a "man/machine" system that had been in operation since 1952 at Central Intelligence Agency (CIA) which, unfortunately, had not been generally known until 1961.[17]

The literature of 1963 indicated an expanding interest

* The author wishes to gratefully acknowledge the generous help given her by International Business Machines Corporation, particularly that of Charles F. Balz, and of Dr. I. A. Warheit and John H. Gustafson.

in KWIC and its variations. New uses for the KWIC technique were tried and reported. Richard L. Storrer described its usefulness in indexing memoranda and letters.[18] It was used for section indexes in cumulative indexes to computer program abstracts, indexes to special collections of publications, indexes to research and development projects, indexes to programs of professional meetings, and in concordances. In other applications, the KWIC technique was used for an index to program titles, indexes to branch office manuals, and an index to manager responsibilities.

Logically, great interest began to be felt and expressed on the subject of titles. Mary Jane Ruhl,[19] Lawrence Papier,[20] Walter Brandenberg,[21] and Jessie Bernard and Charles W. Shilling[22] were among those commenting on the validity of using titles as a basis for indexing and suggesting improvements to titles.

Numerous variations to KWIC had appeared by 1963. E. A. Ripperger and others reported an index called WADEX which combined author entries with word entries.[23] H. R. Newbaker and T. R. Savage wrote on the SWIFT program, which combines features of the KWIC with traditional format, and which depends on titles elaborated to the point that they are called notations of content (NOC).[24] Physindex, a subject index halfway between KWIC indexes and conventional alphabetical indexing, was described by Nicole Chonex, Andre Chonex, and Jean Iung.[25] Need for author participation in writing informative titles and the possible difficulties arising with this need were topics treated by Saul Herner,[26] by T. F. Conolly,[27] and by R. A. Kennedy.[28]

The use of editing to improve the KWIC index was discussed by Phyllis V. Parkins,[29] and pre-editing was mentioned by Robert R. Freeman and G. Malcolm Dyson in a history of the development of *Chemical Titles*.[30] Variations from the Luhn code were noted by Freeman,[31] and by H. East and others.[32]

In 1964 B. B. Lane published the results of a study of title validity in technical and non-technical fields, indicating that in non-technical fields titles reveal the contents less frequently than in technical fields.[33] John M. Sedano in a similar study proposed that the concept of "a technical field" should not be restricted to science or engineering, but should apply equally to any specialized area of knowledge, and cited the excellent use of highly descriptive titles in the Public Affairs and Information Service *Bulletin*.[34] Marguerite Fischer suggested that titles in the non-technical fields might be made KWICable by the stylized use of part titles, a device common to literature of the 17th and 18th centuries.[35]

Without attempting to cover all publications in 1965, it may be noted that a state-of-the-art report by M. E. Stevens on automatic indexing, contains valuable references to KWIC,[36] particularly in the area of early, unpublished materials and work.

● A Preliminary Look at the KWIC Index

The KWIC index and other permuted indexes are among the group of new indexes that are called "unconventional indexes" to differentiate them from subject-heading indexes or classed indexes, which are called "conventional indexes." The underlying principle of the KWIC index is that words instead of concepts can be used for indexing. Keywords — i.e., catchwords or essential words — can be extracted from the title, abstract, or text, and can be used effectively in the index. The context about a keyword helps to define or explain its use, in order to lead the index user to the exact article, paper, or other bit of information he desires. The KWIC index is used chiefly with titles; however, it also can be used with abstracts or with whole texts. Furthermore, it can be edited manually with addition or deletion of words.

Generally, each KWIC index is preceded by an introduction and a stoplist. The stoplist lists words that are not meaningful for indexing purposes and that are excluded from the indexing process. Words included in stoplists vary from index to index and even from time to time in the same index as experience and circumstances dictate.

In the body of the index, each line consists of three parts: the code, the index word, and the context. The parts and their arrangement vary from index to index.

● A History of the KWIC Index

Seen historically, the keyword and the permuted index were not altogether new when Luhn invented the KWIC index, but were devices recovered for machine adaptation from practices of old European libraries.[38] "Indexing by key words, with meaning clarified by context was not new. Scholarly concordances have been known for centuries."[30] A. Crestadero's *Art of Making Catalogues of Libraries*, published in London in 1856, more than one hundred years before the invention of KWIC, referred to the concept of the permutation index.[37] Also, German libraries were using the *schlagwort* — the "catchword," or the "keyword," in the idiom of this paper — in their cataloging procedures one hundred years ago or earlier.[47]

Observing Luhn's background, while keeping in mind the historical precedents for permutation and keyword indexes, it is interesting to speculate that Luhn's early acquaintance with German libraries, as a student or as the son of a German printer, may have led to his later use of the catchword or the keyword in the title as an indexing device for KWIC.

In the early 1950's many people began to look at computers or machines as possible indexing tools. The Central Intelligence Agency as early as 1953 began to prepare a permuted title word index using keypunch, reproducing punch, sorter, and tabulator.[17]

Many different points may be thought of as the beginning of the KWIC index. Black selected the point at which

> the Pontifical Faculty of Philosophy in Milan decided that they would make an analytical index and concordance to the Summa Theologica of St. Thomas Aquinas, and approached IBM about the possibility of having the operations performed on Data Processing machinery. . . . Experience gained in this project contributed towards the development of the KWIC Index.[16]

In 1958 "KWIC was coined by H. P. Luhn . . . at about the same time Citron, Hart, and Ohlman were developing a Permutation Index to the Preprints of the International Conference on Scientific Information. . ." [37] Luhn's "KWIC method offered easy and extremely rapid handling of large volumes of information, relatively simple preparation of the input to the computer, and output of a product which was readily reproduced by photographic offset methods and easy to use." [29]

In the fall of 1958 the Chemical Abstracts Service became convinced that the KWIC index designed by Luhn could be developed as a scheme for indexing the titles of chemical communications.[30] A $150,000 grant by the National Science Foundation's Office of Science Information allowed the Chemical Abstracts Service to develop the keyword indexing scheme, and in April 1960 the Service distributed the first seven thousand sample copies of the index to registrants at the Cleveland meeting of the American Chemical Society.

From 1960 to 1962 over thirty applications of the fundamental techniques of the KWIC concept were made[13] and since 1962 applications have increased even more rapidly.

• Users of the KWIC Concept

It is impossible to compile a complete list of the users of the KWIC concept since it is impossible to determine exactly who is using it. The concept is so simple that anyone with access to a computer can use it easily and effectively for the solution of their particular indexing problems. Even manual use of the concept is possible, although this is practical only for small indexes.[39] Still, consideration of some of the principal users of the KWIC concept and of the manner in which they use the index will indicate its growing importance and popularity.

The Bell Telephone Laboratories, in 1959, decided to use a permuted index. The studies started at that time resulted in the development of a variation of KWIC, the principal characteristic of which is the use of 120 characters per line instead of the 60 characters per line usually used with KWIC indexes.

Biological Abstracts first published its permuted-title subject index in October 1961. The editors christened the index BASIC, standing for Biological Abstracts Subjects in Context. BASIC differed from the earliest KWIC indexes by providing access to abstracts rather than to citations alone. With time, other departures were made from the original KWIC index, and editing, or "vocabulary management," evolved.[29]

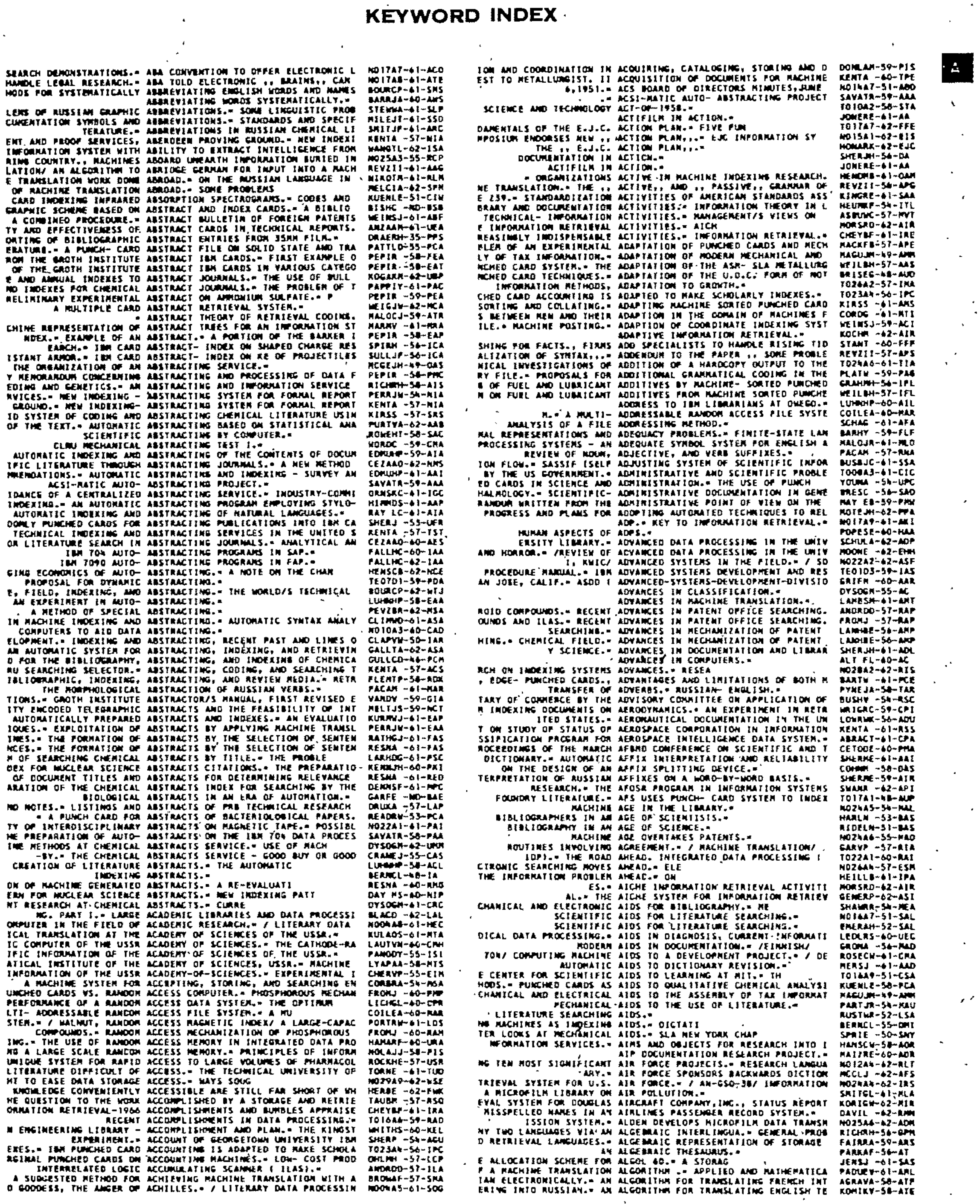
Other users of the KWIC index may be briefly noted: the *KWIC Index* to the *Science Abstracts of China*, issued December 1960 by the MIT Libraries, listed some 3,300 Communist Chinese papers; the *KWIC Index to Neurochemistry*, prepared in 1961 by the Mimosa Frenk Foundation for Applied Neurochemistry in cooperation with IBM, listed some 2,100 papers; *Dissertations in Physics*, an indexed bibliography of the 8,418 doctoral theses accepted by American universities from 1861 to 1959, compiled by the IBM San Jose Research Laboratory and published by Stanford University Press, 1961; *Keywords Index to U. S. Government Technical Reports*, a temporary publication published biweekly by the U. S. Department of Commerce, Office of Technical Services; *Index to Legal Theses and Research Projects*, published in July 1962 by the American Bar Foundation; *Current State Legislation*, a KWIC index of bills enacted by 50 state legislatures, also published by the American Bar Foundation; *Current Medical Terminology*, published in 1964 by the American Medical Association; *Kansas Slavic Index*, published by the University of Kansas; and *Meteorological and Geoastrophysical Titles*, published by the American Meteorological Society. *Chemical Patents*, a publication of the American Chemical Society, was first published in 1960, but failed. A commercial service to libraries, Librarymaster Services, incorporates keyword-in-context indexing among its other services. Companies using KWIC indexes for internal reports or manuals include IBM, Lockheed, the Allison Division of General Motors, and Trans-Canada Airlines. Lawrence Radiation Laboratory and Sandia Laboratories also use KWIC. The Oak Ridge National Laboratory has used it experimentally in a publication that supplements *Nuclear Science Abstracts*, as *Chemical Titles* supplements *Chemical Abstracts*.

Considering, then, the very recent origin of the KWIC concept, its use has spread very rapidly.

• Improvements to KWIC: Variations and Suggestions

1. *Manipulation of the index line.*

There seems to be a general lack of satisfaction with the index line. Users have manipulated it so that many variations have been used. A glance at the first KWIC index reveals excessive white space in the index. When reading across a short line to the associated code number, it is sometimes difficult to determine which code is associated with which title. The difficulty is alleviated by the use of the wrap-around, or recirculated, title. The wrap-around title not only reduces the white space and provides improved readability, it brings more context, or information, onto the index line. White space still exists but not excessively, as seen in Fig. 1.



KEYWORD INDEX

A

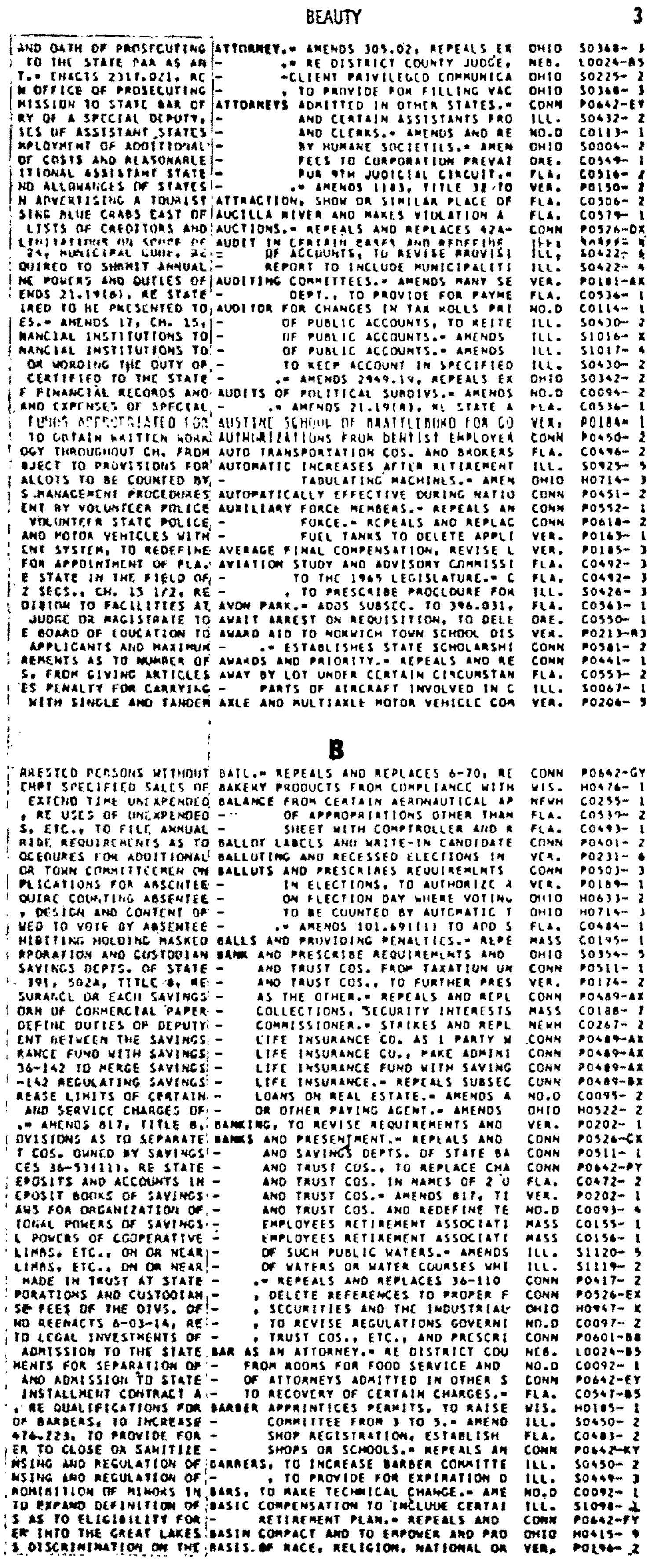
```
SEARCH DEMONSTRATIONS.= ABA CONVENTION TO OFFER ELECTRONIC L  NO17A7-61-ACO
HANDLE LEGAL RESEARCH.= ABA TOLD ELECTRONIC ,, BRAINS,, CAN  NO17A8-61-ATE
HODS FOR SYSTEMATICALLY ABBREVIATING ENGLISH WORDS AND NAMES  BOURCP-61-SMS
                        ABBREVIATING WORDS SYSTEMATICALLY.=  BARRJA-60-AWS
LEMS OF RUSSIAN GRAPHIC ABBREVIATIONS.= SOME LINGUISTIC PROB  STEWWA-61-SLP
CUMENTATION SYMBOLS AND ABBREVIATIONS.= STANDARDS AND SPECIF  MILEJT-61-SSD
             TERATURE.= ABBREVIATIONS IN RUSSIAN CHEMICAL LI  SMITJF-61-ARC
ENT AND PROOF SERVICES, ABERDEEN PROVING GROUND.= NEW INDEXI  KENTA -57-NIA
INFORMATION SYSTEM WITH ABILITY TO EXTRACT INTELLIGENCE FROM  WANGTL-62-ISA
RING COUNTRY., MACHINES ABOARD UNEARTH INFORMATION BURIED IN  NO25A3-55-MCP
LATION/ AN ALGORITHM TO ABRIDGE GERMAN FOR INPUT INTO A MACH  REVZII-61-AAG
E TRANSLATION WORK DONE ABROAD.= ON THE RUSSIAN LANGUAGE IN  MIROTM-61-RLM
 OF MACHINE TRANSLATION ABROAD.= SOME PROBLEMS  MELCIA-62-SPM
 CARD INDEXING INFRARED ABSORPTION SPECTROGRAMS.= CODES AND  KUENLE-51-CIM
GRAPHIC SCHEME BASED ON ABSTRACT AND INDEX CARDS.= A BIBLIO  BISHC -ND-BSB
 A COMBINED PROCEDURE.= ABSTRACT BULLETIN OF FOREIGN PATENTS  WEINSJ-61-ABF
TY AND EFFECTIVENESS OF ABSTRACT CARDS IN TECHNICAL REPORTS.  ANZAAM-61-UEA
ORTING OF BIBLIOGRAPHIC ABSTRACT ENTRIES FROM 35MM FILM.=  DRAERH-35-PPS
ERATURE.= A PUNCH- CARD ABSTRACT FILE ON SOLID STATE AND TRA  PATTLD-55-PCA
ROM THE GROTH INSTITUTE ABSTRACT IBM CARDS.= FIRST EXAMPLE O  PEPIR -58-FEA
 OF THE GROTH INSTITUTE ABSTRACT IBM CARDS IN VARIOUS CATEGO  PEPIR -58-EAT
E AND ANNUAL INDEXES TO ABSTRACT JOURNALS.= THE USE OF BULL  KOGAKM-62-UBP
ND INDEXES FOR CHEMICAL ABSTRACT JOURNALS.= THE PROBLEM OF T  PAPPIY-61-PAC
RELIMINARY EXPERIMENTAL ABSTRACT ON AMMONIUM SULFATE.= P  PEPIR -59-PEA
        A MULTIPLE CARD ABSTRACT RETRIEVAL SYSTEM.=  WEIGJW-62-MCA
                      = ABSTRACT THEORY OF RETRIEVAL CODING.  MALOCJ-59-ATR
CHINE REPRESENTATION OF ABSTRACT TREES FOR AN INFORMATION ST  MARNV -61-MRA
   NDEX.= EXAMPLE OF AN ABSTRACT.= A PORTION OF THE BARKER I  PEPIR -58-EAP
       EARCH.= IBM CARD ABSTRACT- INDEX ON SHAPED CHARGE RES  SPIRM -56-ICA
ISTANT ARMOR.= IBM CARD ABSTRACT- INDEX ON KE OF PROJECTILES  SULLJF-56-ICA
 THE ORGANIZATION OF AN ABSTRACTING SERVICE.=  MCGEJM-49-OAS
Y MEMORANDUM CONCERNING ABSTRACTING AND PROCESSING OF DATA F  PEPIR -58-PMC
EDING AND GENETICS.= AN ABSTRACTING AND INFORMATION SERVICE  RICHMH-58-AIS
RVICES.= NEW INDEXING - ABSTRACTING SYSTEM FOR FORMAL REPORT  PERRJW-54-NIA
 GROUND.= NEW INDEXING- ABSTRACTING SYSTEM FOR FORMAL REPORT  KENTA -57-NIA
ID SYSTEM OF CODING AND ABSTRACTING CHEMICAL LITERATURE USIN  KIRSS -57-SRS
OF THE TEXT.= AUTOMATIC ABSTRACTING BASED ON STATISTICAL ANA  PURTVA-62-AAB
             SCIENTIFIC ABSTRACTING BY COMPUTER.=  ROMENT-58-SAC
        CLRU MECHANICAL ABSTRACTING TEST I.=  WORDC -59-CMA
 AUTOMATIC INDEXING AND ABSTRACTING OF THE CONTENTS OF DOCUM  EDMUHP-59-AIA
IFIC LITERATURE THROUGH ABSTRACTING JOURNALS.= A NEW METHOD  CEZAAO-62-NMS
MMENDATIONS.= AUTOMATIC ABSTRACTING AND INDEXING - SURVEY AN  EDMUHP-61-AAI
       ACSI-MATIC AUTO- ABSTRACTING PROJECT.=  SAVATR-59-AAA
IDANCE OF A CENTRALIZED ABSTRACTING SERVICE.= INDUSTRY-COMMI  ORNSKC-61-IGC
INDEXING.= AN AUTOMATIC ABSTRACTING PROGRAM EMPLOYING STYLO-  HIMMDS-61-AAP
 AUTOMATIC INDEXING AND ABSTRACTING OF NATURAL LANGUAGES.=  RAY LC-61-AIA
DOMLY PUNCHED CARDS FOR ABSTRACTING PUBLICATIONS INTO IBM CA  SHERJ -55-UFR
 TECHNICAL INDEXING AND ABSTRACTING SERVICES IN THE UNITED S  KENTA -57-TST
OR LITERATURE SEARCH IN ABSTRACTING JOURNALS.= ANALYTICAL AN  CEZAAO-60-AES
          IBM 704 AUTO- ABSTRACTING PROGRAMS IN SAP.=  FALLHC-60-IAA
         IBM 7090 AUTO- ABSTRACTING PROGRAMS IN FAP.=  FALLHC-62-IAA
GING ECONOMICS OF AUTO- ABSTRACTING.= A NOTE ON THE CHAN  HENSCB-62-NCE
   PROPOSAL FOR DYNAMIC ABSTRACTING.=  TEO7D1-59-PDA
E, FIELD, INDEXING, AND ABSTRACTING.= THE WORLD/S TECHNICAL  BOURCP-62-WTJ
 AN EXPERIMENT IN AUTO- ABSTRACTING.=  LUHNHP-58-EAA
  . A METHOD OF SPECIAL ABSTRACTING.=  PEVZBR-62-MSA
IN MACHINE INDEXING AND ABSTRACTING.= AUTOMATIC SYNTAX ANALY  CLIMMO-61-ASA
  COMPUTERS TO AID DATA ABSTRACTING.=  NO10A3-60-CAD
ELOPMENT.= INDEXING AND ABSTRACTING, RECENT PAST AND LINES O  CLAPVW-50-IAR
AN AUTOMATIC SYSTEM FOR ABSTRACTING, INDEXING, AND RETRIEVIN  GALLTA-62-ASA
D FOR THE BIBLIOGRAPHY, ABSTRACTING, AND INDEXING OF CHEMICA  GULLCD-46-PCM
RU SEARCHING SELECTOR.= ABSTRACTING, CODING, AND SEARCHING T  KENTA -57-ACS
IBLIOGRAPHIC, INDEXING, ABSTRACTING, AND REVIEW MEDIA.= RETR  FLEMTP-58-RDX
      THE MORPHOLOGICAL ABSTRACTION OF RUSSIAN VERBS.=  FACAM -61-MAR
TIONS.= GROTH INSTITUTE ABSTRACTOR/S MANUAL, FIRST REVISED E  VANDV -59-GIA
ITY ENCODED TELEGRAPHIC ABSTRACTS AND THE FEASIBILITY OF INT  MELTJS-59-NCT
 AUTOMATICALLY PREPARED ABSTRACTS AND INDEXES.= AN EVALUATIO  KURMWJ-61-EAP
IQUES.= EXPLOITATION OF ABSTRACTS BY APPLYING MACHINE TRANSL  PERRJW-61-EAA
INES.= THE FORMATION OF ABSTRACTS BY THE SELECTION OF SENTEN  RATHGJ-61-FAS
NCES.= THE FORMATION OF ABSTRACTS BY THE SELECTION OF SENTEN  RESNA -61-FAS
M OF SEARCHING CHEMICAL ABSTRACTS BY TITLE.= THE PROBLE  LARHDG-61-PSC
DEX FOR NUCLEAR SCIENCE ABSTRACTS CITATIONS.= THE PREPARATIO  KENNJH-60-PKI
 OF DOCUMENT TITLES AND ABSTRACTS FOR DETERMINING RELEVANCE  RESNA -61-RED
ARATION OF THE CHEMICAL ABSTRACTS INDEX FOR SEARCHING BY THE  DENNSF-61-MPC
             BIOLOGICAL ABSTRACTS IN AN ERA OF AUTOMATION.=  GARFE -ND-BAE
NO NOTES.= LISTINGS AND ABSTRACTS OF PRB TECHNICAL RESEARCH  DRUKA -57-LAP
     = A PUNCH CARD FOR ABSTRACTS OF BACTERIOLOGICAL PAPERS.  READRW-53-PCA
TY OF INTERDISCIPLINARY ABSTRACTS ON MAGNETIC TAPE.= POSSIBL  NO22A1-61-PAI
HE PREPARATION OF AUTO- ABSTRACTS ON THE IBM 704 DATA PROCES  SAVATR-58-PAA
INE METHODS AT CHEMICAL ABSTRACTS SERVICE.= USE OF MACH  DYSOGH-62-UMM
     -BY.= THE CHEMICAL ABSTRACTS SERVICE - GOOD BUY OR GOOD  CRANEJ-55-CAS
 CREATION OF LITERATURE ABSTRACTS.= THE AUTOMATIC  LUHNHP-58-ACL
               INDEXING ABSTRACTS.=  BERNCL-48-IA
ON OF MACHINE GENERATED ABSTRACTS.= A RE-EVALUATI  RESNA -60-RMG
ERN FOR NUCLEAR SCIENCE ABSTRACTS.= NEW INDEXING PATT  DAY MS-60-NIP
NT RESEARCH AT CHEMICAL ABSTRACTS.= CURRE  DYSOGH-61-CRC
     NG. PART I.= LARGE ACADEMIC LIBRARIES AND DATA PROCESSI  BLACD -62-LAL
OMPUTER IN THE FIELD OF ACADEMIC RESEARCH.= / LITERARY DATA  MOONAB-61-HEC
ICAL TRANSLATION AT THE ACADEMY OF SCIENCES OF THE USSR.=  KULAOS-61-MTA
IC COMPUTER OF THE USSR ACADEMY OF SCIENCES.= THE CATHODE-RA  LAUTVN-60-CMH
IFIC INFORMATION OF THE ACADEMY OF SCIENCES OF THE USSR.=  PANODY-55-ISI
ATICAL INSTITUTE OF THE ACADEMY OF SCIENCES, USSR.= MACHINE  LYAPAA-58-MTS
INFORMATION OF THE USSR ACADEMY-OF-SCIENCES.= EXPERIMENTAL I  CHERVP-55-EIM
   A MACHINE SYSTEM FOR ACCEPTING, STORING, AND SEARCHING EN  CORBRA-54-MSA
UNCHED CARDS VS. RANDOM ACCESS COMPUTER.= PHOSPHOROUS MECHAN  FROMJ -60-PMP
PERFORMANCE OF A RANDOM ACCESS DATA SYSTEM.= THE OPTIMUM  LICHGL-60-OPR
LTI- ADDRESSABLE RANDOM ACCESS FILE SYSTEM.= A MU  COILEA-60-MAR
STEM.= / WALNUT, RANDOM ACCESS MAGNETIC INDEX/ A LARGE-CAPAC  PORTRW-61-LDS
     COMPOUNDS.= RANDOM ACCESS MECHANIZATION OF PHOSPHOROUS  FROMJ -60-RAM
ING.= THE USE OF RANDOM ACCESS MEMORY IN INTEGRATED DATA PRO  HAMARF-60-URA
NG A LARGE SCALE RANDOM ACCESS MEMORY.= PRINCIPLES OF INFORM  HOLAJJ-58-PIS
UNIQUE SYSTEM FOR RAPID ACCESS TO LARGE VOLUMES OF PHARMACOL  ROCKHE-57-USR
LITERATURE DIFFICULT OF ACCESS.= THE TECHNICAL UNIVERSITY OF  TORNE -61-TUD
NT TO EASE DATA STORAGE ACCESS.= WAYS SOUG  NO29A9-62-WSE
 KNOWLEDGE CONVENIENTLY ACCESSIBLE ARE STILL FAR SHORT OF WH  HERBE -62-FWK
HE QUESTION TO THE WORK ACCOMPLISHED BY A STORAGE AND RETRIE  TAUBM -57-RSQ
ORMATION RETRIEVAL-1966 ACCOMPLISHMENTS AND BUMBLES APPRAISE  CHEYBF-61-IRA
                 RECENT ACCOMPLISHMENTS IN DATA PROCESSING.=  TO16AA-59-RAD
M ENGINEERING LIBRARY - ACCOMPLISHMENT AND PLAN.= THE KINOST  WHITHS-60-KEL
           EXPERIMENT.= ACCOUNT OF GEORGETOWN UNIVERSITY IBM  SHERP -54-AGU
EXES.= IBM PUNCHED CARD ACCOUNTING IS ADAPTED TO MAKE SCHOLA  TO23AN-56-IPC
RGINAL PUNCHED CARDS ON ACCOUNTING MACHINES.= LOW- COST PROD  OHLMH -57-LCP
     INTERRELATED LOGIC ACCUMULATING SCANNER ( ILAS).=  ANDRDD-57-ILA
 A SUGGESTED METHOD FOR ACHIEVING MACHINE TRANSLATION WITH A  BROWAF-57-SMA
O GODDESS, THE ANGER OF ACHILLES.= / LITERARY DATA PROCESSIN  MOONAB-61-SOG
```

```
ION AND COORDINATION IN ACQUIRING, CATALOGING, STORING AND D  DONLAM-59-PIS
EST TO METALLURGIST. II ACQUISITION OF DOCUMENTS FOR MACHINE  KENTA -60-TPE
               6,1951.= ACS BOARD OF DIRECTORS MINUTES,JUNE  NO14A7-51-ABD
                     .= ACSI-MATIC AUTO- ABSTRACTING PROJECT  SAVATR-59-AAA
 SCIENCE AND TECHNOLOGY ACT-OF-1958.=  TO10A2-58-STA
                        ACTIFILM IN ACTION.=  JONERE-61-AA
DAMENTALS OF THE E.J.C. ACTION PLAN.= FIVE FUN  TO17A7-62-FFE
MPOSIUM ENDORSES NEW ,, ACTION PLAN,,.= EJC INFORMATION SY  NO15A1-62-EIS
          THE ,, E.J.C. ACTION PLAN,,.=  HOWARK-62-EJC
       DOCUMENTATION IN ACTION.=  SHERJH-56-DA
            ACTIFILM IN ACTION.=  JONERE-61-AA
        = ORGANIZATIONS ACTIVE IN MACHINE INDEXING RESEARCH.  HENDMB-61-OAM
NE TRANSLATION.= THE ,, ACTIVE,, AND ,, PASSIVE,, GRAMMAR OF  REVZII-56-APG
E 239.= STANDARDIZATION ACTIVITIES OF AMERICAN STANDARDS ASS  KINGRE-61-SAA
BRARY AND DOCUMENTATION ACTIVITIES.= INFORMATION THEORY IN L  HEUMKF-54-ITL
 TECHNICAL- INFORMATION ACTIVITIES.= MANAGEMENT/S VIEWS ON  ASMUWC-57-MVT
F INFORMATION RETRIEVAL ACTIVITIES.= AICH  MORSRD-62-AIR
REASINGLY INDISPENSABLE ACTIVITIES.= INFORMATION RETRIEVAL.=  CHEYBF-61-IRE
BLEM OF AN EXPERIMENTAL ADAPTATION OF PUNCHED CARDS AND MECH  MACKFB-57-APE
LY OF TAX INFORMATION.= ADAPTATION OF MODERN MECHANICAL AND  MAGUJM-49-AMM
NCHED CARD SYSTEM.= THE ADAPTATION OF THE ASM- SLA METALLURG  WEILBH-57-AAS
NCHED CARD TECHNIQUES.= ADAPTATION OF THE U.D.C. FORM OF NOT  BRISEG-48-AUD
   INFORMATION METHODS, ADAPTATION TO GROWTH.=  TO26A2-57-IMA
CHED CARD ACCOUNTING IS ADAPTED TO MAKE SCHOLARLY INDEXES.=  TO23AN-56-IPC
SORTING AND COLLATING.= ADAPTING MACHINE SORTED PUNCHED CARD  KIRSS -61-AMS
S BETWEEN MEN AND THEIR ADAPTION IN THE DOMAIN OF MACHINES F  CORDG -61-MTI
ILE.= MACHINE POSTING.= ADAPTION OF COORDINATE INDEXING SYST  WEINSJ-59-ACI
                        ADAPTIVE INFORMATION RETRIEVAL.=  KOCHM -62-AIR
SHING FOR FACTS., FIRMS ADD SPECIALISTS TO HANDLE RISING TID  STANT -60-FFF
ALIZATION OF SYNTAX,,.= ADDENDUM TO THE PAPER ,, SOME PROBLE  REVZII-57-APS
NICAL INVESTIGATIONS OF ADDITION OF A HARDCOPY OUTPUT TO THE  TO24A6-61-TIA
RY FILE.= PROPOSALS FOR ADDITIONAL GRAMMATICAL CODING IN THE  PLATW -59-PAG
R OF FUEL AND LUBRICANT ADDITIVES BY MACHINE- SORTED PUNCHED  GRAHMH-56-IFL
N ON FUEL AND LUBRICANT ADDITIVES FROM MACHINE SORTED PUNCHE  WEILBH-57-IFL
                        ADDRESS TO IBM LIBRARIANS AT OWEGO.=  LUHNHP-60-AIL
           M.= A MULTI- ADDRESSABLE RANDOM ACCESS FILE SYSTE  COILEA-60-MAR
     ANALYSIS OF A FILE ADDRESSING METHOD.=  SCHAG -61-AFA
MAL REPRESENTATIONS AND ADEQUACY PROBLEMS.= FINITE-STATE LAN  BARHY -59-FLF
PROCESSING SYSTEMS - AN ADEQUATE SYMBOL SYSTEM FOR ENGLISH A  MALOJR-61-MLO
        REVIEW OF NOUN, ADJECTIVE, AND VERB SUFFIXES.=  PACAM -57-RNA
ION FLOW.= SASSIF (SELF ADJUSTING SYSTEM OF SCIENTIFIC INFOR  BUSBJC-61-SSA
 BY THE US GOVERNMENT.= ADMINISTRATIVE AND SCIENTIFIC PROBLE  TO08A3-61-CIC
ED CARDS IN SCIENCE AND ADMINISTRATION.= THE USE OF PUNCH  YOUNA -54-UPC
HALMOLOGY.= SCIENTIFIC- ADMINISTRATIVE DOCUMENTATION IN GENE  WRESC -56-SAD
RANDUM WRITTEN FROM THE ADMINISTRATIVE POINT OF VIEW ON THE  MAY EB-59-PMW
 PROGRESS AND PLANS FOR ADOPTING AUTOMATED TECHNIQUES TO REL  MOTEJH-62-PPA
                        ADP.= KEY TO INFORMATION RETRIEVAL.=  NO17A9-61-AKI
       HUMAN ASPECTS OF ADPS.=  POPESE-60-HAA
       ERSITY LIBRARY.= ADVANCED DATA PROCESSING IN THE UNIV  SCHULA-62-ADP
AND HORROR.= /REVIEW OF ADVANCED DATA PROCESSING IN THE UNIV  MOONE -62-EHH
               I, KWIC/ ADVANCED SYSTEMS IN THE FIELD.= / SD  NO22A2-62-ASF
 PROCEDURE MANUAL.= IBM ADVANCED SYSTEMS DEVELOPMENT AND RES  TEO1D3-59-IAS
AN JOSE, CALIF.= ASDD ( ADVANCED-SYSTEMS-DEVELOPMENT-DIVISIO  GRIFM -60-AAR
                        ADVANCES IN CLASSIFICATION.=  DYSOGH-55-AC
                        ADVANCES IN MACHINE TRANSLATION.=.  LAMBSM-61-AMT
ROID COMPOUNDS.= RECENT ADVANCES IN PATENT OFFICE SEARCHING.  ANDRDD-57-RAP
OUNDS AND ILAS.= RECENT ADVANCES IN PATENT OFFICE SEARCHING.  FROMJ -57-RAP
            SEARCHING.= ADVANCES IN MECHANIZATION OF PATENT  LANHBE-56-AMP
HING.= CHEMICAL FIELD.= ADVANCES IN MECHANIZATION OF PATENT  LANHBE-56-AMP
            Y SCIENCE.= ADVANCES IN DOCUMENTATION AND LIBRAR  SHERJH-61-ADL
                        ADVANCES IN COMPUTERS.=  ALT FL-60-AC
RCH ON INDEXING SYSTEMS ADVANCES.= RESEA  NO28A2-62-RIS
, EDGE- PUNCHED CARDS., ADVANTAGES AND LIMITATIONS OF BOTH M  BARTW -61-MCE
            TRANSFER OF ADVERBS.= RUSSIAN- ENGLISH.=  PYNEJA-58-TAR
TARY OF COMMERCE BY THE ADVISORY COMMITTEE ON APPLICATION OF  BUSHV -54-RSC
M INDEXING DOCUMENTS ON AERODYNAMICS.= AN EXPERIMENT IN RETR  WRIGRC-59-CPI
          ITED STATES.= AERONAUTICAL DOCUMENTATION IN THE UN  LOWRWK-56-ADU
T ON STUDY OF STATUS OF AEROSPACE CORPORATION IN INFORMATION  KENTA -61-RSS
SSIFICATION PROGRAM FOR AEROSPACE INTELLIGENCE DATA SYSTEM.=  ABRACT-61-CPA
ROCEEDINGS OF THE MARCH AFBMD CONFERENCE ON SCIENTIFIC AND T  CETODE-60-PMA
 DICTIONARY.= AUTOMATIC AFFIX INTERPRETATION AND RELIABILITY  SHERME-61-AAI
    ON THE DESIGN OF AN AFFIX SPLITTING DEVICE.=  COHNM -58-DAS
TERPRETATION OF RUSSIAN AFFIXES ON A WORD-BY-WORD BASIS.=  SHERME-59-AIR
         RESEARCH.= THE AFOSR PROGRAM IN INFORMATION SYSTEMS  SWANR -62-API
   FOUNDRY LITERATURE.= AFS USES PUNCH- CARD SYSTEM TO INDEX  TO17A1-48-AUP
                MACHINE AGE IN THE LIBRARY.=  NO24A5-54-MAL
   BIBLIOGRAPHERS IN AN AGE OF SCIENTISTS.=  HARLN -53-BAS
     BIBLIOGRAPHY IN AN AGE OF SCIENCE.=  RIDELN-51-BAS
                MACHINE AGE OVERTAKES PATENTS.=  NO24A6-55-MAO
     ROUTINES INVOLVING AGREEMENT.= / MACHINE TRANSLATION/  GARVP -57-RIA
        IDP).= THE ROAD AHEAD. INTEGRATED DATA PROCESSING (  TO22A1-60-RAI
CTRONIC SEARCHING MOVES AHEAD.= ELE  NO26A4-57-ESM
THE INFORMATION PROBLEM AHEAD.= ON  HEILLB-61-IPA
                   ES.= AICHE INFORMATION RETRIEVAL ACTIVITI  MORSRD-62-AIR
               AL.= THE AICHE SYSTEM FOR INFORMATION RETRIEV  GENERP-62-ASI
CHANICAL AND ELECTRONIC AIDS FOR BIBLIOGRAPHY.= ME  SHAWRR-54-MEA
             SCIENTIFIC AIDS FOR LITERATURE SEARCHING.=  NO16A7-51-SAL
             SCIENTIFIC AIDS FOR LITERATURE SEARCHING.=  EMERAH-52-SAL
DICAL DATA PROCESSING.= AIDS IN DIAGNOSIS, CURRENT INFORMATI  LEDLRS-60-UEC
                 MODERN AIDS IN DOCUMENTATION.= /FINNISH/  GRONA -56-MAD
 TON/ COMPUTING MACHINE AIDS TO A DEVELOPMENT PROJECT.= / DE  ROSECH-61-CMA
              AUTOMATIC AIDS TO DICTIONARY REVISION.=  MERSJ -61-AAD
E CENTER FOR SCIENTIFIC AIDS TO LEARNING AT MIT.= TH  TO16A9-51-CSA
HODS.= PUNCHED CARDS AS AIDS TO QUALITATIVE CHEMICAL ANALYSI  KUENLE-58-PCA
CHANICAL AND ELECTRICAL AIDS TO THE ASSEMBLY OF TAX INFORMAT  MAGUJM-49-AMM
             MECHANICAL AIDS TO THE USE OF LITERATURE.=  PARTJR-54-MAU
   LITERATURE SEARCHING AIDS.=  RUSTWR-52-LSA
NG MACHINES AS INDEXING AIDS.= DICTATI  BERNCL-55-DMI
TER LOOKS AT MECHANICAL AIDS.= SLA NEW YORK CHAP  SPRIE -50-SNY
  NFORMATION SERVICES.= AIMS AND OBJECTS FOR RESEARCH INTO I  HANSCW-58-AOR
                        AIP DOCUMENTATION RESEARCH PROJECT.=  MAIZRE-60-ADR
NG TEN MOST SIGNIFICANT AIR FORCE PROJECTS.= RESEARCH LANGUA  NO12A4-62-RLT
                  ARY.= AIR FORCE SPONSORS BACKWARDS DICTION  MCCLJ -62-AFS
TRIEVAL SYSTEM FOR U.S. AIR FORCE.= / AN-GSQ-38/ INFORMATION  NO24A4-62-IRS
 A MICROFILM LIBRARY ON AIR POLLUTION.=  SMITGL-61-MLA
EVAL SYSTEM FOR DOUGLAS AIRCRAFT COMPANY,INC., STATUS REPORT  KORIGM-62-MIR
 MISSPELLED NAMES IN AN AIRLINES PASSENGER RECORD SYSTEM.=  DAVIL -62-RMM
        ISSION SYSTEM.= ALDEN DEVELOPS MICROFILM DATA TRANSM  NO25A6-62-ADM
NY TWO LANGUAGES VIA AN ALGEBRAIC INTERLINGUA.= GENERAL PROB  RICHRH-56-GPM
D RETRIEVAL LANGUAGES.= ALGEBRAIC REPRESENTATION OF STORAGE  FAIRRA-59-ARS
                     AN ALGEBRAIC THESAURUS.=  PARKAF-56-AT
E ALLOCATION SCHEME FOR ALGOL 60.= A STORAG  JENSJ -61-SAS
F A MACHINE TRANSLATION ALGORITHM .= APPLIED AND MATHEMATICA  PADUEV-61-AML
IAN ELECTRONICALLY.= AN ALGORITHM FOR TRANSLATING FRENCH INT  AGRAVA-58-ATF
ERING INTO RUSSIAN.= AN ALGORITHM FOR TRANSLATING ENGLISH TE  KOMIKV-58-ATE
```

FIG. 1. Sample from Index with Recirculated Title

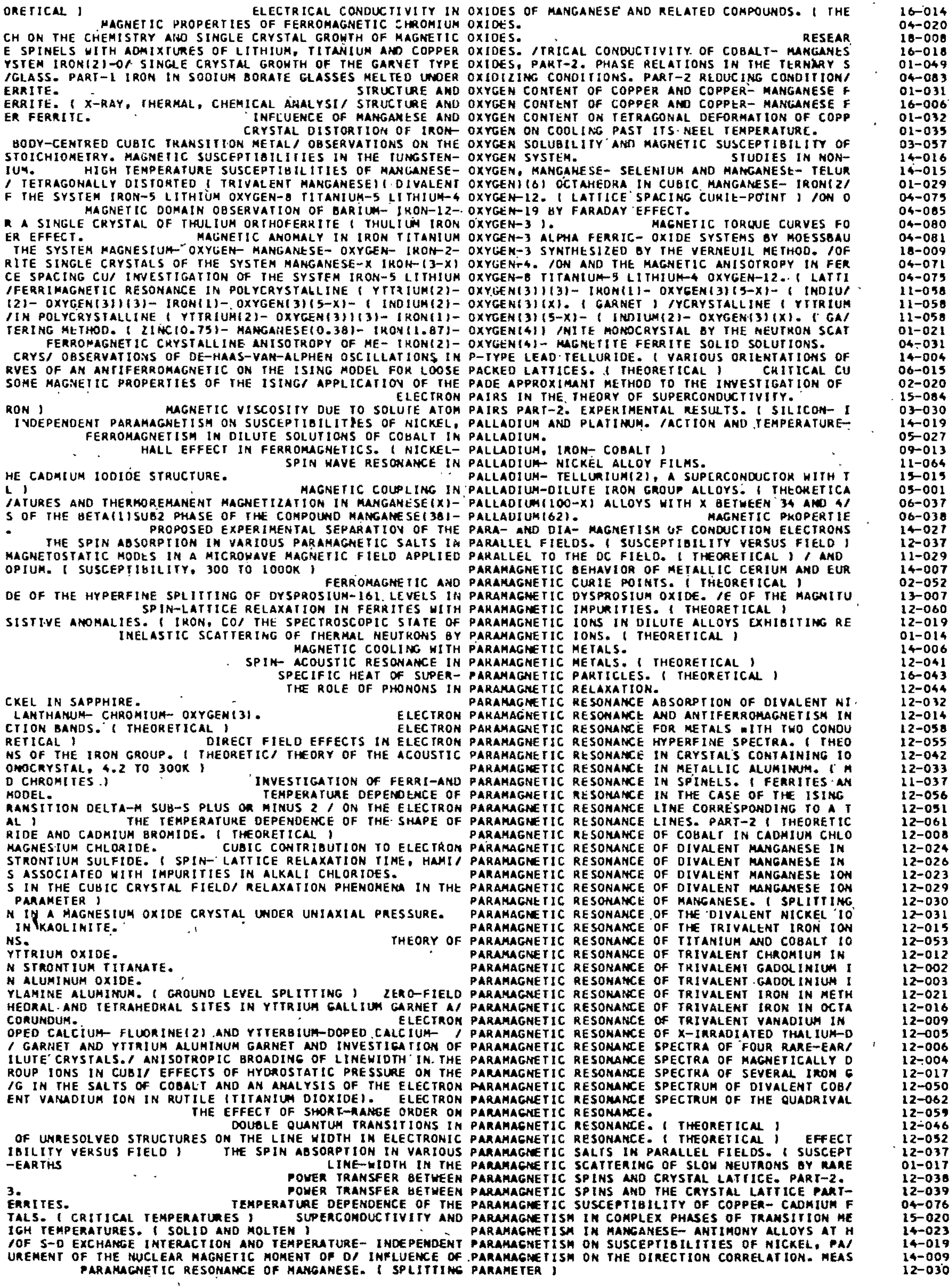
The editors of *Biological Abstracts* and the editors of *Current State Legislation* have darkened the column of title fragments to the left of the column of index words in an attempt to improve readability. An example from *Current State Legislation* is shown in Fig. 2.

The wish to retain the full title has led to Bell Telephone's use of a 120-character index line instead of the 60-character index line normally used in KWIC. The advantage claimed for the long line is that only 2% of the titles listed with the line are chopped off, whereas 30% of the titles listed with a 60-character line are chopped off. The problem of excessive white space appears with the use of the 120-character line, but does not seem to be as distracting as it is with the 60-character line. An example of the Bell Telephone Index is provided in Fig. 3.



```
                                    BEAUTY                                3

AND OATH OF PROSECUTING ATTORNEY.* AMENDS 305.02, REPEALS EX  OHIO  S0368- 3
 TO THE STATE BAR AS AN -       .* RE DISTRICT COUNTY JUDGE,  NEB.  L0024-B5
T.* ENACTS 2317.021, RC -       -CLIENT PRIVILEGED COMMUNICA  OHIO  S0225- 2
N OFFICE OF PROSECUTING -       , TO PROVIDE FOR FILLING VAC  OHIO  S0368- 3
MISSION TO STATE BAR OF ATTORNEYS ADMITTED IN OTHER STATES.*  CONN  P0642-EY
RY OF A SPECIAL DEPUTY, -         AND CERTAIN ASSISTANTS PRO  ILL.  S0432- 2
ICS OF ASSISTANT STATES -         AND CLERKS.* AMENDS AND RE  NO.D  C0113- 1
MPLOYMENT OF ADDITIONAL -         BY HUMANE SOCIETIES.* AMEN  OHIO  S0004- 2
OF COSTS AND REASONABLE -         FEES TO CORPORATION PREVAI  ORE.  C0549- 1
ITIONAL ASSISTANT STATE -         FOR 9TH JUDICIAL CIRCUIT.*  FLA.  C0516- 2
ND ALLOWANCES OF STATES -       .* AMENDS 1183, TITLE 32 TO   VER.  P0150- 2
N ADVERTISING A TOURIST ATTRACTION, SHOW OR SIMILAR PLACE OF  FLA.  C0506- 2
SING BLUE CRABS EAST OF AUCILLA RIVER AND MAKES VIOLATION A   FLA.  C0579- 1
 LISTS OF CREDITORS AND AUCTIONS.* REPEALS AND REPLACES 42A-  CONN  P0576-DX
LIMITATIONS ON SCOPE OF AUDIT IN CERTAIN CASES AND REDEFINE   ILL.  S0422- 4
 24, MUNICIPAL CODE, RE -     OF ACCOUNTS, TO REVISE PROVISI  ILL.  S0422- 4
QUIRED TO SUBMIT ANNUAL -     REPORT TO INCLUDE MUNICIPALITI  ILL.  S0422- 4
NE POWERS AND DUTIES OF AUDITING COMMITTEES.* AMENDS MANY SE  VER.  P0181-AX
ENDS 21.19(6), RE STATE -         DEPT., TO PROVIDE FOR PAYME FLA.  C0536- 1
IRED TO BE PRESENTED TO AUDITOR FOR CHANGES IN TAX ROLLS PRI  NO.D  C0114- 1
ES.* AMENDS 17, CH. 15, -       OF PUBLIC ACCOUNTS, TO REITE  ILL.  S0430- 2
NANCIAL INSTITUTIONS TO -       OF PUBLIC ACCOUNTS.* AMENDS   ILL.  S1016- X
NANCIAL INSTITUTIONS TO -       OF PUBLIC ACCOUNTS.* AMENDS   ILL.  S1017- 4
 OR WORDING THE DUTY OF -       TO KEEP ACCOUNT IN SPECIFIED  ILL.  S0430- 2
 CERTIFIED TO THE STATE -     .* AMENDS 2949.19, REPEALS EX   OHIO  S0342- 2
F FINANCIAL RECORDS AND AUDITS OF POLITICAL SUBDIVS.* AMENDS  NO.D  C0094- 2
AND EXPENSES OF SPECIAL -    .* AMENDS 21.19(8), RE STATE A   FLA.  C0536- 1
 FUNDS APPROPRIATED FOR AUSTINE SCHOOL OF BRATTLEBORO FOR CO  VER.  P0184- 1
 TO OBTAIN WRITTEN NORA AUTHORIZATIONS FROM DENTIST EMPLOYER  CONN  P0450- 2
OGY THROUGHOUT CH. FROM AUTO TRANSPORTATION COS. AND BROKERS  FLA.  C0496- 2
BJECT TO PROVISIONS FOR AUTOMATIC INCREASES AFTER RETIREMENT  ILL.  S0925- 5
ALLOTS TO BE COUNTED BY -          TABULATING MACHINES.* AMEN OHIO  H0714- 3
S MANAGEMENT PROCEDURES AUTOMATICALLY EFFECTIVE DURING NATIO  CONN  P0451- 2
ENT BY VOLUNTEER POLICE AUXILIARY FORCE MEMBERS.* REPEALS AN  CONN  P0552- 1
 VOLUNTEER STATE POLICE -          FORCE.* REPEALS AND REPLAC CONN  P0618- 2
AND MOTOR VEHICLES WITH -          FUEL TANKS TO DELETE APPLI VER.  P0163- 1
ENT SYSTEM, TO REDEFINE AVERAGE FINAL COMPENSATION, REVISE L  VER.  P0185- 3
FOR APPOINTMENT OF FLA. AVIATION STUDY AND ADVISORY COMMISSI  FLA.  C0492- 3
E STATE IN THE FIELD OF -          TO THE 1965 LEGISLATURE.* C FLA. C0492- 3
Z SECS., CH. 15 1/2, RE -       , TO PRESCRIBE PROCEDURE FOR  ILL.  S0426- 3
DIVISION TO FACILITIES AT AVON PARK.* ADDS SUBSEC. TO 396.031, FLA. C0563- 1
 JUDGE OR MAGISTRATE TO AWAIT ARREST ON REQUISITION, TO DELE  ORE.  C0550- 1
E BOARD OF EDUCATION TO AWARD AID TO NORWICH TOWN SCHOOL DIS  VER.  P0213-R3
 APPLICANTS AND MAXIMUM -     .* ESTABLISHES STATE SCHOLARSHI CONN  P0581- 2
REMENTS AS TO NUMBER OF AWARDS AND PRIORITY.* REPEALS AND RE  CONN  P0441- 1
S. FROM GIVING ARTICLES AWAY BY LOT UNDER CERTAIN CIRCUMSTAN  FLA.  C0553- 2
ES PENALTY FOR CARRYING -    PARTS OF AIRCRAFT INVOLVED IN C  ILL.  S0067- 1
 WITH SINGLE AND TANDEM AXLE AND MULTIAXLE MOTOR VEHICLE COM  VER.  P0206- 5

                                      B

ARRESTED PERSONS WITHOUT BAIL.* REPEALS AND REPLACES 6-70, RE CONN  P0642-GY
EMPT SPECIFIED SALES OF BAKERY PRODUCTS FROM COMPLIANCE WITH  WIS.  H0476- 1
 EXTEND TIME UNEXPENDED BALANCE FROM CERTAIN AERONAUTICAL AP  NEWH  C0255- 1
, RE USES OF UNEXPENDED -         OF APPROPRIATIONS OTHER THAN FLA. C0539- 2
S. ETC., TO FILE ANNUAL -         SHEET WITH COMPTROLLER AND R FLA. C0493- 1
RIDE REQUIREMENTS AS TO BALLOT LABELS AND WRITE-IN CANDIDATE  CONN  P0401- 2
OCEDURES FOR ADDITIONAL BALLOTING AND RECESSED ELECTIONS IN   VER.  P0231- 6
OR TOWN COMMITTEEMEN ON BALLOTS AND PRESCRIBES REQUIREMENTS   CONN  P0503- 3
PLICATIONS FOR ABSENTEE -         IN ELECTIONS, TO AUTHORIZE A VER. P0169- 1
QUIRE COUNTING ABSENTEE -         ON ELECTION DAY WHERE VOTING OHIO H0633- 2
, DESIGN AND CONTENT OF -         TO BE COUNTED BY AUTOMATIC T OHIO H0714- 3
WED TO VOTE BY ABSENTEE -       .* AMENDS 101.691(1) TO ADD S FLA.  C0484- 1
HIBITING HOLDING MASKED BALLS AND PROVIDING PENALTIES.* REPE  MASS  C0195- 1
RPORATION AND CUSTODIAN BANK AND PRESCRIBE REQUIREMENTS AND   OHIO  S0354- 5
SAVINGS DEPTS. OF STATE -    AND TRUST COS. FROM TAXATION UN  CONN  P0511- 1
 391, 502A, TITLE 8, RE -    AND TRUST COS., TO FURTHER PRES  VER.  P0174- 2
SURANCE OR EACH SAVINGS -    AS THE OTHER.* REPEALS AND REPL  CONN  P0489-AX
ORM OF COMMERCIAL PAPER -    COLLECTIONS, SECURITY INTERESTS  MASS  C0188- 7
DEFINE DUTIES OF DEPUTY -    COMMISSIONER.* STRIKES AND REPL  NEWH  C0267- 2
ENT BETWEEN THE SAVINGS -    LIFE INSURANCE CO. AS 1 PARTY W  CONN  P0489-AX
RANCE FUND WITH SAVINGS -    LIFE INSURANCE CO., MAKE ADMINI  CONN  P0489-AX
36-142 TO MERGE SAVINGS -    LIFE INSURANCE FUND WITH SAVING  CONN  P0489-AX
-142 REGULATING SAVINGS -    LIFE INSURANCE.* REPEALS SUBSEC  CONN  P0489-BX
REASE LIMITS OF CERTAIN -    LOANS ON REAL ESTATE.* AMENDS A  NO.D  C0095- 2
 AND SERVICE CHARGES OF -    OR OTHER PAYING AGENT.* AMENDS   OHIO  H0522- 2
.* AMENDS 817, TITLE 8, BANKING, TO REVISE REQUIREMENTS AND   VER.  P0202- 1
DVISIONS AS TO SEPARATE BANKS AND PRESENTMENT.* REPEALS AND   CONN  P0526-CX
T COS. OWNED BY SAVINGS -     AND SAVINGS DEPTS. OF STATE BA  CONN  P0511- 1
CES 36-53(11), RE STATE -     AND TRUST COS., TO REPLACE CHA  CONN  P0642-PY
EPOSITS AND ACCOUNTS IN -     AND TRUST COS. IN NAMES OF 2 U  FLA.  C0472- 2
EPOSIT BOOKS OF SAVINGS -     AND TRUST COS.* AMENDS 817, TI  VER.  P0202- 1
AWS FOR ORGANIZATION OF -     AND TRUST COS. AND REDEFINE TE  NO.D  C0093- 4
IONAL POWERS OF SAVINGS -     EMPLOYEES RETIREMENT ASSOCIATI  MASS  C0155- 1
L POWERS OF COOPERATIVE -     EMPLOYEES RETIREMENT ASSOCIATI  MASS  C0156- 1
LIMBS, ETC., ON OR NEAR -     OF SUCH PUBLIC WATERS.* AMENDS  ILL.  S1120- 5
LIMBS, ETC., ON OR NEAR -     OF WATERS OR WATER COURSES WHI  ILL.  S1119- 2
 MADE IN TRUST AT STATE -   .* REPEALS AND REPLACES 36-110   CONN  P0417- 2
PORATIONS AND CUSTODIAN -    , DELETE REFERENCES TO PROPER F  CONN  P0526-EX
SE FEES OF THE DIVS. OF -    , SECURITIES AND THE INDUSTRIAL  OHIO  H0947- X
ND REENACTS 6-03-14, RE -    , TO REVISE REGULATIONS GOVERNI  NO.D  C0097- 2
TO LEGAL INVESTMENTS OF -    , TRUST COS., ETC., AND PRESCRI  CONN  P0601-BB
 ADMISSION TO THE STATE BAR AS AN ATTORNEY.* RE DISTRICT COU  NEB.  L0024-B5
MENTS FOR SEPARATION OF -   FROM ROOMS FOR FOOD SERVICE AND   NO.D  C0092- 1
 AND ADMISSION TO STATE -   OF ATTORNEYS ADMITTED IN OTHER S  CONN  P0642-EY
 INSTALLMENT CONTRACT A -   TO RECOVERY OF CERTAIN CHARGES.*  FLA.  C0547-B5
, RE QUALIFICATIONS FOR BARBER APPRINTICES PERMITS, TO RAISE  WIS.  H0185- 1
OF BARBERS, TO INCREASE -       COMMITTEE FROM 3 TO 5.* AMEND ILL.  S0450- 2
474.223, TO PROVIDE FOR -       SHOP REGISTRATION, ESTABLISH  FLA.  C0483- 2
ER TO CLOSE OR SANITIZE -       SHOPS OR SCHOOLS.* REPEALS AN CONN  P0642-KY
NSING AND REGULATION OF BARBERS, TO INCREASE BARBER COMMITTE  ILL.  S0450- 2
NSING AND REGULATION OF -     , TO PROVIDE FOR EXPIRATION O   ILL.  S0449- 3
ROHIBITION OF MINORS IN BARS, TO MAKE TECHNICAL CHANGE.* AME  NO.D  C0092- 1
TO EXPAND DEFINITION OF BASIC COMPENSATION TO INCLUDE CERTAI  ILL.  S1098- 1
S AS TO ELIGIBILITY FOR -     RETIREMENT PLAN.* REPEALS AND   CONN  P0642-FY
ER INTO THE GREAT LAKES BASIN COMPACT AND TO EMPOWER AND PRO  OHIO  H0415- 9
S DISCRIMINATION ON THE BASIS OF RACE, RELIGION, NATIONAL OR  VER.  P0196- 2
```

Fig. 2. Sample from *Current State Legislation:* An Index Using Darkened Title Fragments



```
ORETICAL )                     ELECTRICAL CONDUCTIVITY IN OXIDES OF MANGANESE AND RELATED COMPOUNDS. ( THE        16-014
           MAGNETIC PROPERTIES OF FERROMAGNETIC CHROMIUM OXIDES.                                                 04-020
CH ON THE CHEMISTRY AND SINGLE CRYSTAL GROWTH OF MAGNETIC OXIDES.                                        RESEAR  18-008
E SPINELS WITH ADMIXTURES OF LITHIUM, TITANIUM AND COPPER OXIDES. /TRICAL CONDUCTIVITY OF COBALT- MANGANES        16-018
YSTEM IRON(2)-O/ SINGLE CRYSTAL GROWTH OF THE GARNET TYPE OXIDES, PART-2. PHASE RELATIONS IN THE TERNARY S        01-049
/GLASS. PART-1 IRON IN SODIUM BORATE GLASSES MELTED UNDER OXIDIZING CONDITIONS. PART-2 REDUCING CONDITION/        04-083
ERRITE.                                     STRUCTURE AND OXYGEN CONTENT OF COPPER AND COPPER- MANGANESE F        01-031
ERRITE. ( X-RAY, THERMAL, CHEMICAL ANALYSI/ STRUCTURE AND OXYGEN CONTENT OF COPPER AND COPPER- MANGANESE F        16-006
ER FERRITE.                     INFLUENCE OF MANGANESE AND OXYGEN CONTENT ON TETRAGONAL DEFORMATION OF COPP       01-032
                            CRYSTAL DISTORTION OF IRON- OXYGEN ON COOLING PAST ITS NEEL TEMPERATURE.              01-035
 BODY-CENTRED CUBIC TRANSITION METAL/ OBSERVATIONS ON THE OXYGEN SOLUBILITY AND MAGNETIC SUSCEPTIBILITY OF        03-057
STOICHIOMETRY. MAGNETIC SUSCEPTIBILITIES IN THE TUNGSTEN- OXYGEN SYSTEM.                      STUDIES IN NON-     14-016
IUM.      HIGH TEMPERATURE SUSCEPTIBILITIES OF MANGANESE- OXYGEN, MANGANESE- SELENIUM AND MANGANESE- TELUR        14-015
/ TETRAGONALLY DISTORTED ( TRIVALENT MANGANESE)( DIVALENT OXYGEN)(6) OCTAHEDRA IN CUBIC MANGANESE- IRON(2/        01-029
F THE SYSTEM IRON-5 LITHIUM OXYGEN-8 TITANIUM-5 LITHIUM-4 OXYGEN-12. ( LATTICE SPACING CURIE-POINT ) /ON O        04-075
         MAGNETIC DOMAIN OBSERVATION OF BARIUM- IRON-12- OXYGEN-19 BY FARADAY EFFECT.                             04-085
R A SINGLE CRYSTAL OF THULIUM ORTHOFERRITE ( THULIUM IRON OXYGEN-3 ).                MAGNETIC TORQUE CURVES FO    04-080
ER EFFECT.                MAGNETIC ANOMALY IN IRON TITANIUM OXYGEN-3 ALPHA FERRIC- OXIDE SYSTEMS BY MOESSBAU      04-081
 THE SYSTEM MAGNESIUM- OXYGEN- MANGANESE- OXYGEN- IRON-2- OXYGEN-3 SYNTHESIZED BY THE VERNEUIL METHOD. /OF        18-009
RITE SINGLE CRYSTALS OF THE SYSTEM MANGANESE-X IRON-(3-X) OXYGEN-4. /ON AND THE MAGNETIC ANISOTROPY IN FER        04-071
CE SPACING CU/ INVESTIGATION OF THE SYSTEM IRON-5 LITHIUM OXYGEN-8 TITANIUM-5 LITHIUM-4 OXYGEN-12. ( LATTI        04-075
/FERRIMAGNETIC RESONANCE IN POLYCRYSTALLINE ( YTTRIUM(2)- OXYGEN(3))(3)- IRON(1)- OXYGEN(3)(5-X)- ( INDIU/        11-058
(2)- OXYGEN(3))(3)- IRON(1)- OXYGEN(3)(5-X)- ( INDIUM(2)- OXYGEN(3)(X). ( GARNET ) /YCRYSTALLINE ( YTTRIUM        11-058
/IN POLYCRYSTALLINE ( YTTRIUM(2)- OXYGEN(3))(3)- IRON(1)- OXYGEN(3)(5-X)- ( INDIUM(2)- OXYGEN(3)(X). ( GA/        11-058
TERING METHOD. ( ZINC(0.75)- MANGANESE(0.38)- IRON(1.87)- OXYGEN(4)) /NITE MONOCRYSTAL BY THE NEUTRON SCAT        01-021
     FERROMAGNETIC CRYSTALLINE ANISOTROPY OF ME- IRON(2)- OXYGEN(4)- MAGNETITE FERRITE SOLID SOLUTIONS.           04-031
 CRYS/ OBSERVATIONS OF DE-HAAS-VAN-ALPHEN OSCILLATIONS IN P-TYPE LEAD TELLURIDE. ( VARIOUS ORIENTATIONS OF        14-004
RVES OF AN ANTIFERROMAGNETIC ON THE ISING MODEL FOR LOOSE PACKED LATTICES. ( THEORETICAL )      CRITICAL CU       06-015
SOME MAGNETIC PROPERTIES OF THE ISING/ APPLICATION OF THE PADE APPROXIMANT METHOD TO THE INVESTIGATION OF         02-020
                                                  ELECTRON PAIRS IN THE THEORY OF SUPERCONDUCTIVITY.              15-084
RON )             MAGNETIC VISCOSITY DUE TO SOLUTE ATOM PAIRS PART-2. EXPERIMENTAL RESULTS. ( SILICON- I          03-030
 INDEPENDENT PARAMAGNETISM ON SUSCEPTIBILITIES OF NICKEL, PALLADIUM AND PLATINUM. /ACTION AND TEMPERATURE-        14-019
         FERROMAGNETISM IN DILUTE SOLUTIONS OF COBALT IN PALLADIUM.                                               05-027
                   HALL EFFECT IN FERROMAGNETICS. ( NICKEL- PALLADIUM, IRON- COBALT )                             09-013
                                  SPIN WAVE RESONANCE IN PALLADIUM- NICKEL ALLOY FILMS.                           11-064
HE CADMIUM IODIDE STRUCTURE.                              PALLADIUM- TELLURIUM(2), A SUPERCONDUCTOR WITH T        15-015
L )                                   MAGNETIC COUPLING IN PALLADIUM-DILUTE IRON GROUP ALLOYS. ( THEORETICA       05-001
/ATURES AND THERMOREMANENT MAGNETIZATION IN MANGANESE(X)- PALLADIUM(100-X) ALLOYS WITH X BETWEEN 34 AND 4/        06-037
S OF THE BETA(1)SUB2 PHASE OF THE COMPOUND MANGANESE(38)- PALLADIUM(62).                   MAGNETIC PROPERTIE     06-038
.                   PROPOSED EXPERIMENTAL SEPARATION OF THE PARA- AND DIA- MAGNETISM OF CONDUCTION ELECTRONS       14-027
     THE SPIN ABSORPTION IN VARIOUS PARAMAGNETIC SALTS IN PARALLEL FIELDS. ( SUSCEPTIBILITY VERSUS FIELD )        12-037
MAGNETOSTATIC MODES IN A MICROWAVE MAGNETIC FIELD APPLIED PARALLEL TO THE DC FIELD. ( THEORETICAL ) / AND         11-029
OPIUM. ( SUSCEPTIBILITY, 300 TO 1000K )                   PARAMAGNETIC BEHAVIOR OF METALLIC CERIUM AND EUR        14-007
                                      FERROMAGNETIC AND PARAMAGNETIC CURIE POINTS. ( THEORETICAL )                02-052
DE OF THE HYPERFINE SPLITTING OF DYSPROSIUM-161 LEVELS IN PARAMAGNETIC DYSPROSIUM OXIDE. /E OF THE MAGNITU        13-007
                 SPIN-LATTICE RELAXATION IN FERRITES WITH PARAMAGNETIC IMPURITIES. ( THEORETICAL )                12-060
SISTIVE ANOMALIES. ( IRON, CO/ THE SPECTROSCOPIC STATE OF PARAMAGNETIC IONS IN DILUTE ALLOYS EXHIBITING RE        12-019
              INELASTIC SCATTERING OF THERMAL NEUTRONS BY PARAMAGNETIC IONS. ( THEORETICAL )                      01-014
                                    MAGNETIC COOLING WITH PARAMAGNETIC METALS.                                    14-006
                             SPIN- ACOUSTIC RESONANCE IN PARAMAGNETIC METALS. ( THEORETICAL )                     12-041
                                  SPECIFIC HEAT OF SUPER- PARAMAGNETIC PARTICLES. ( THEORETICAL )                 16-043
                                   THE ROLE OF PHONONS IN PARAMAGNETIC RELAXATION.                                12-044
CKEL IN SAPPHIRE.                                         PARAMAGNETIC RESONANCE ABSORPTION OF DIVALENT NI        12-032
 LANTHANUM- CHROMIUM- OXYGEN(3).                 ELECTRON PARAMAGNETIC RESONANCE AND ANTIFERROMAGNETISM IN        12-014
CTION BANDS. ( THEORETICAL )                     ELECTRON PARAMAGNETIC RESONANCE FOR METALS WITH TWO CONDU        12-058
RETICAL )               DIRECT FIELD EFFECTS IN ELECTRON PARAMAGNETIC RESONANCE HYPERFINE SPECTRA. ( THEO         12-055
NS OF THE IRON GROUP. ( THEORETIC/ THEORY OF THE ACOUSTIC PARAMAGNETIC RESONANCE IN CRYSTALS CONTAINING IO        12-042
ONOCRYSTAL, 4.2 TO 300K )                                 PARAMAGNETIC RESONANCE IN METALLIC ALUMINUM. ( M        12-033
D CHROMITES .)                  INVESTIGATION OF FERRI-AND PARAMAGNETIC RESONANCE IN SPINELS. ( FERRITES AN       11-037
MODEL.                       TEMPERATURE DEPENDENCE OF PARAMAGNETIC RESONANCE IN THE CASE OF THE ISING            12-056
RANSITION DELTA-M SUB-S PLUS OR MINUS 2 / ON THE ELECTRON PARAMAGNETIC RESONANCE LINE CORRESPONDING TO A T        12-051
AL )            THE TEMPERATURE DEPENDENCE OF THE SHAPE OF PARAMAGNETIC RESONANCE LINES. PART-2 ( THEORETIC       12-061
RIDE AND CADMIUM BROMIDE. ( THEORETICAL )                 PARAMAGNETIC RESONANCE OF COBALT IN CADMIUM CHLO        12-008
MAGNESIUM CHLORIDE.        CUBIC CONTRIBUTION TO ELECTRON PARAMAGNETIC RESONANCE OF DIVALENT MANGANESE IN         12-024
STRONTIUM SULFIDE. ( SPIN- LATTICE RELAXATION TIME, HAMI/ PARAMAGNETIC RESONANCE OF DIVALENT MANGANESE IN         12-026
S ASSOCIATED WITH IMPURITIES IN ALKALI CHLORIDES.         PARAMAGNETIC RESONANCE OF DIVALENT MANGANESE ION        12-023
S IN THE CUBIC CRYSTAL FIELD/ RELAXATION PHENOMENA IN THE PARAMAGNETIC RESONANCE OF DIVALENT MANGANESE ION        12-029
 PARAMETER )                                              PARAMAGNETIC RESONANCE OF MANGANESE. ( SPLITTING        12-030
N IN A MAGNESIUM OXIDE CRYSTAL UNDER UNIAXIAL PRESSURE.   PARAMAGNETIC RESONANCE OF THE DIVALENT NICKEL IO        12-031
 IN KAOLINITE.                                            PARAMAGNETIC RESONANCE OF THE TRIVALENT IRON ION        12-015
NS.                                             THEORY OF PARAMAGNETIC RESONANCE OF TITANIUM AND COBALT IO        12-053
YTTRIUM OXIDE.                                            PARAMAGNETIC RESONANCE OF TRIVALENT CHROMIUM IN         12-012
N STRONTIUM TITANATE.                                     PARAMAGNETIC RESONANCE OF TRIVALENT GADOLINIUM I        12-002
N ALUMINUM OXIDE.                                         PARAMAGNETIC RESONANCE OF TRIVALENT GADOLINIUM I        12-003
YLAMINE ALUMINUM. ( GROUND LEVEL SPLITTING )   ZERO-FIELD PARAMAGNETIC RESONANCE OF TRIVALENT IRON IN METH        12-021
HEDRAL AND TETRAHEDRAL SITES IN YTTRIUM GALLIUM GARNET A/ PARAMAGNETIC RESONANCE OF TRIVALENT IRON IN OCTA        12-016
CORUNDUM.                                        ELECTRON PARAMAGNETIC RESONANCE OF TRIVALENT VANADIUM IN         12-009
OPED CALCIUM- FLUORINE(2) AND YTTERBIUM-DOPED CALCIUM-  / PARAMAGNETIC RESONANCE OF X-IRRADIATED THALIUM-D        12-005
/ GARNET AND YTTRIUM ALUMINUM GARNET AND INVESTIGATION OF PARAMAGNETIC RESONANCE SPECTRA OF FOUR RARE-EAR/        12-006
ILUTE CRYSTALS./ ANISOTROPIC BROADING OF LINEWIDTH IN THE PARAMAGNETIC RESONANCE SPECTRA OF MAGNETICALLY D        12-004
ROUP IONS IN CUBI/ EFFECTS OF HYDROSTATIC PRESSURE ON THE PARAMAGNETIC RESONANCE SPECTRA OF SEVERAL IRON G        12-017
/G IN THE SALTS OF COBALT AND AN ANALYSIS OF THE ELECTRON PARAMAGNETIC RESONANCE SPECTRUM OF DIVALENT COB/        12-050
ENT VANADIUM ION IN RUTILE (TITANIUM DIOXIDE).   ELECTRON PARAMAGNETIC RESONANCE SPECTRUM OF THE QUADRIVAL        12-062
                       THE EFFECT OF SHORT-RANGE ORDER ON PARAMAGNETIC RESONANCE.                                 12-059
                            DOUBLE QUANTUM TRANSITIONS IN PARAMAGNETIC RESONANCE. ( THEORETICAL )                 12-046
 OF UNRESOLVED STRUCTURES ON THE LINE WIDTH IN ELECTRONIC PARAMAGNETIC RESONANCE. ( THEORETICAL )    EFFECT       12-052
IBILITY VERSUS FIELD )    THE SPIN ABSORPTION IN VARIOUS PARAMAGNETIC SALTS IN PARALLEL FIELDS. ( SUSCEPT         12-037
-EARTHS                                    LINE-WIDTH IN THE PARAMAGNETIC SCATTERING OF SLOW NEUTRONS BY RARE     01-017
                                POWER TRANSFER BETWEEN PARAMAGNETIC SPINS AND CRYSTAL LATTICE. PART-2.            12-038
3.                              POWER TRANSFER BETWEEN PARAMAGNETIC SPINS AND THE CRYSTAL LATTICE PART-           12-039
ERRITES.                      TEMPERATURE DEPENDENCE OF THE PARAMAGNETIC SUSCEPTIBILITY OF COPPER- CADMIUM F      04-076
TALS. ( CRITICAL TEMPERATURES )    SUPERCONDUCTIVITY AND PARAMAGNETISM IN COMPLEX PHASES OF TRANSITION ME         15-020
IGH TEMPERATURES. ( SOLID AND MOLTEN )                    PARAMAGNETISM IN MANGANESE- ANTIMONY ALLOYS AT H        14-023
/OF S-D EXCHANGE INTERACTION AND TEMPERATURE- INDEPENDENT PARAMAGNETISM ON SUSCEPTIBILITIES OF NICKEL, PA/        14-019
UREMENT OF THE NUCLEAR MAGNETIC MOMENT OF D/ INFLUENCE OF PARAMAGNETISM ON THE DIRECTION CORRELATION. MEAS        14-009
        PARAMAGNETIC RESONANCE OF MANGANESE. ( SPLITTING PARAMETER )                                              12-030
```

FIG. 3. Sample from Bell Telephone Index: An Index Using 120 Characters per Line

Many users seek to increase the effective use of the index by taking the keyword out of context, forming a keyword-out-of-context (KWOC) index. An example of a KWOC index is provided in Fig. 4.

A number of other indexes, although they are not called KWOC indexes, list the keyword out of context. An IBM index similar to the KWOC index is illustrated in Fig. 5.

Wolfe of IBM has produced innovations in programming that provide a keyword-out-of-context index with the full title in its natural order. The index, which is called a KWIC index, is illustrated in Fig. 6.

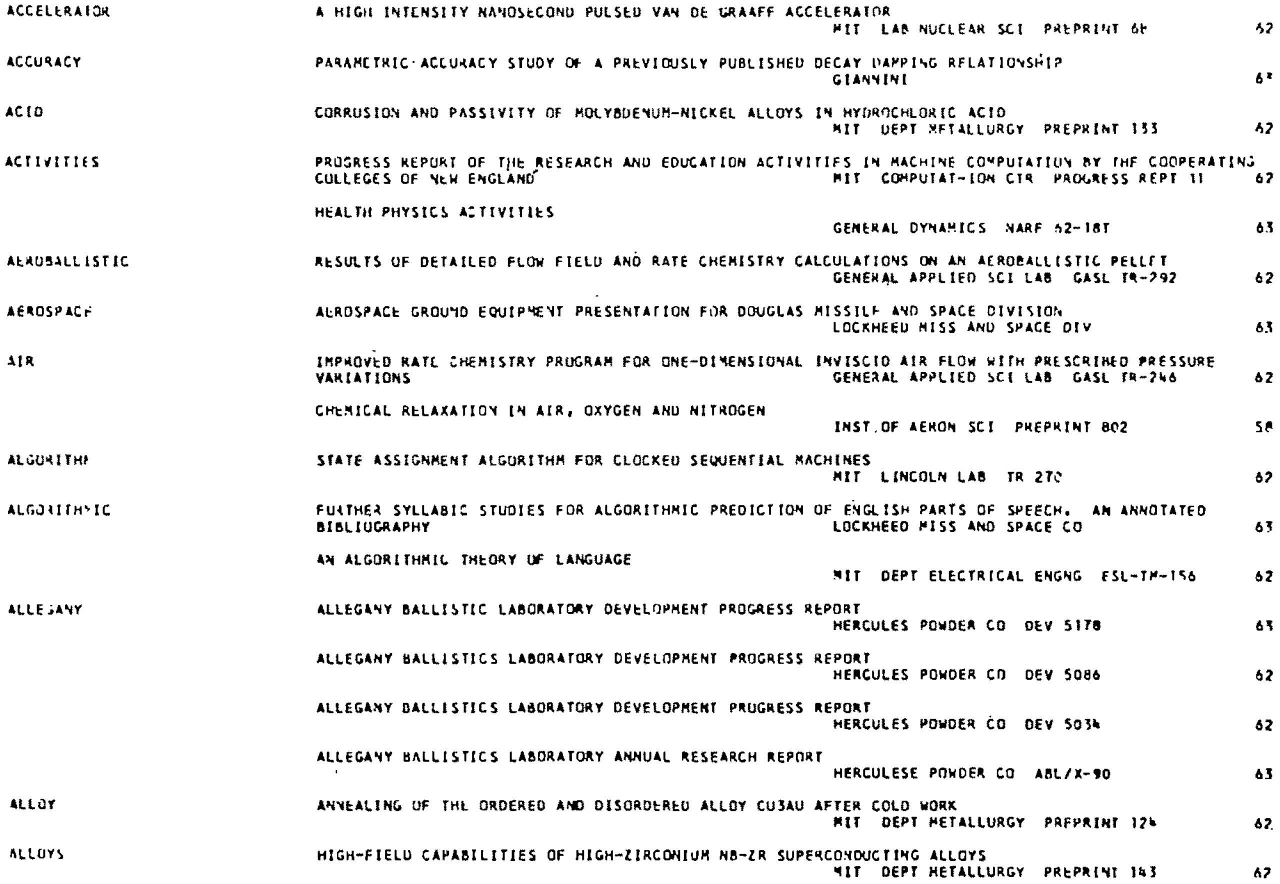
The KWOC indexes are very popular. In addition to the KWOC indexes illustrated above, the following are KWOC indexes: *Keyword Titles,* published by the Office of Technical Services; *Scientific and Technical Aerospace Reports,* published by the National Space Aeronautics and Space Association; and *International Aerospace Abstracts,* published by the American Institute of Aeronautics and Astronautics. However, as Youden states, KWOC indexes "make the search for multiword phrases . . . much more difficult."[40] The index illustrated in Fig. 7 is offered as an improved version of the index illustrated in Fig. 6, differing from it in that it leaves the keyword in context, attempts a full citation, and combines an author index.

2. *Variations of the code.*

The alphanumeric code in the KWIC index line is composed of different elements, according to the special requirements of the index. The Luhn code, the first to be used and the most commonly used,

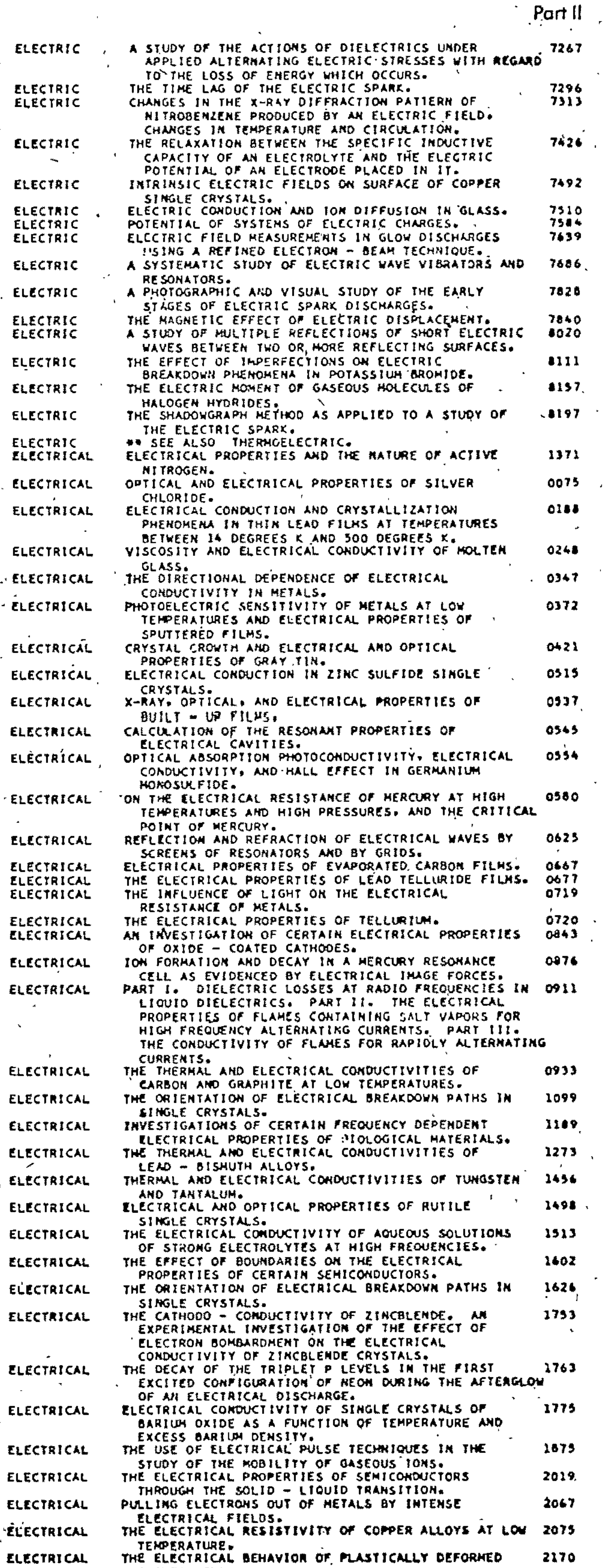
> is derived from factual data inherent in a document as evinced by the publisher's printed identification, comprising the following elements:
>
> 1. The name of the author (or senior author) or originating agency.
> 2. The year of publication.
> 3. The title of the document.
>
> . . . .
>
> The code comprises eleven character positions. The first six are derived from the name of the author or originating agency, the next two consist of the ten's and unit digit of the year of publication, and the last three are derived from the title.
>
> The above code format was chosen over other possible variations for the reason that when bibliographical entries are ordered in alphabetical sequence in accordance with this code, the utility of the resulting listing as an author index is not seriously impaired since the variations between this order and that demanded by the fully spelled words are slight.[1]

An early departure from this code was made by the editors of *Chemical Abstracts* in 1962. Reader criticism of



```
ACCELERATOR      A HIGH INTENSITY NANOSECOND PULSED VAN DE GRAAFF ACCELERATOR
                                                         MIT  LAB NUCLEAR SCI  PREPRINT 68          62

ACCURACY         PARAMETRIC ACCURACY STUDY OF A PREVIOUSLY PUBLISHED DECAY DAMPING RELATIONSHIP
                                                         GIANNINI                                   6*

ACID             CORROSION AND PASSIVITY OF MOLYBDENUM-NICKEL ALLOYS IN HYDROCHLORIC ACID
                                                         MIT  DEPT METALLURGY  PREPRINT 133         62

ACTIVITIES       PROGRESS REPORT OF THE RESEARCH AND EDUCATION ACTIVITIES IN MACHINE COMPUTATION BY THE COOPERATING
                 COLLEGES OF NEW ENGLAND                 MIT  COMPUTAT-ION CTR  PROGRESS REPT 11    62

                 HEALTH PHYSICS ACTIVITIES
                                                         GENERAL DYNAMICS  NARF 62-18T              63

AEROBALLISTIC    RESULTS OF DETAILED FLOW FIELD AND RATE CHEMISTRY CALCULATIONS ON AN AEROBALLISTIC PELLET
                                                         GENERAL APPLIED SCI LAB  GASL TR-292       62

AEROSPACE        AEROSPACE GROUND EQUIPMENT PRESENTATION FOR DOUGLAS MISSILE AND SPACE DIVISION
                                                         LOCKHEED MISS AND SPACE DIV                63

AIR              IMPROVED RATE CHEMISTRY PROGRAM FOR ONE-DIMENSIONAL INVISCID AIR FLOW WITH PRESCRIBED PRESSURE
                 VARIATIONS                              GENERAL APPLIED SCI LAB  GASL TR-246       62

                 CHEMICAL RELAXATION IN AIR, OXYGEN AND NITROGEN
                                                         INST.OF AERON SCI  PREPRINT 802            58

ALGORITHM        STATE ASSIGNMENT ALGORITHM FOR CLOCKED SEQUENTIAL MACHINES
                                                         MIT  LINCOLN LAB  TR 270                   62

ALGORITHMIC      FURTHER SYLLABIC STUDIES FOR ALGORITHMIC PREDICTION OF ENGLISH PARTS OF SPEECH.  AN ANNOTATED
                 BIBLIOGRAPHY                            LOCKHEED MISS AND SPACE CO                 63

                 AN ALGORITHMIC THEORY OF LANGUAGE
                                                         MIT  DEPT ELECTRICAL ENGNG  ESL-TM-156     62

ALLEGANY         ALLEGANY BALLISTIC LABORATORY DEVELOPMENT PROGRESS REPORT
                                                         HERCULES POWDER CO  DEV 5178               63

                 ALLEGANY BALLISTICS LABORATORY DEVELOPMENT PROGRESS REPORT
                                                         HERCULES POWDER CO  DEV 5086               62

                 ALLEGANY BALLISTICS LABORATORY DEVELOPMENT PROGRESS REPORT
                                                         HERCULES POWDER CO  DEV 5034               62

                 ALLEGANY BALLISTICS LABORATORY ANNUAL RESEARCH REPORT
                                                         HERCULESE POWDER CO  ABL/X-90              63

ALLOY            ANNEALING OF THE ORDERED AND DISORDERED ALLOY CU3AU AFTER COLD WORK
                                                         MIT  DEPT METALLURGY  PREPRINT 124         62

ALLOYS           HIGH-FIELD CAPABILITIES OF HIGH-ZIRCONIUM NB-ZR SUPERCONDUCTING ALLOYS
                                                         MIT  DEPT METALLURGY  PREPRINT 143         62
```

FIG. 4. Sample from Douglas Missiles and Space Library Index: A KWOC Index



```
                                                                    Part II
ELECTRIC     A STUDY OF THE ACTIONS OF DIELECTRICS UNDER               7267
               APPLIED ALTERNATING ELECTRIC STRESSES WITH REGARD
               TO THE LOSS OF ENERGY WHICH OCCURS.
ELECTRIC     THE TIME LAG OF THE ELECTRIC SPARK.                       7296
ELECTRIC     CHANGES IN THE X-RAY DIFFRACTION PATTERN OF               7313
               NITROBENZENE PRODUCED BY AN ELECTRIC FIELD.
               CHANGES IN TEMPERATURE AND CIRCULATION.
ELECTRIC     THE RELAXATION BETWEEN THE SPECIFIC INDUCTIVE             7426
               CAPACITY OF AN ELECTROLYTE AND THE ELECTRIC
               POTENTIAL OF AN ELECTRODE PLACED IN IT.
ELECTRIC     INTRINSIC ELECTRIC FIELDS ON SURFACE OF COPPER            7492
               SINGLE CRYSTALS.
ELECTRIC     ELECTRIC CONDUCTION AND ION DIFFUSION IN GLASS.           7510
ELECTRIC     POTENTIAL OF SYSTEMS OF ELECTRIC CHARGES.                 7584
ELECTRIC     ELCCTRIC FIELD MEASUREMENTS IN GLOW DISCHARGES            7639
               USING A REFINED ELECTRON - BEAM TECHNIQUE.
ELECTRIC     A SYSTEMATIC STUDY OF ELECTRIC WAVE VIBRATORS AND         7686
               RESONATORS.
ELECTRIC     A PHOTOGRAPHIC AND VISUAL STUDY OF THE EARLY              7828
               STAGES OF ELECTRIC SPARK DISCHARGES.
ELECTRIC     THE MAGNETIC EFFECT OF ELECTRIC DISPLACEMENT.             7840
ELECTRIC     A STUDY OF MULTIPLE REFLECTIONS OF SHORT ELECTRIC         8020
               WAVES BETWEEN TWO OR MORE REFLECTING SURFACES.
ELECTRIC     THE EFFECT OF IMPERFECTIONS ON ELECTRIC                   8111
               BREAKDOWN PHENOMENA IN POTASSIUM BROMIDE.
ELECTRIC     THE ELECTRIC MOMENT OF GASEOUS MOLECULES OF               8157
               HALOGEN HYDRIDES.
ELECTRIC     THE SHADOWGRAPH METHOD AS APPLIED TO A STUDY OF           8197
               THE ELECTRIC SPARK.
ELECTRIC     ** SEE ALSO  THERMOELECTRIC.
ELECTRICAL   ELECTRICAL PROPERTIES AND THE NATURE OF ACTIVE            1371
               NITROGEN.
ELECTRICAL   OPTICAL AND ELECTRICAL PROPERTIES OF SILVER               0075
               CHLORIDE.
ELECTRICAL   ELECTRICAL CONDUCTION AND CRYSTALLIZATION                 0188
               PHENOMENA IN THIN LEAD FILMS AT TEMPERATURES
               BETWEEN 14 DEGREES K AND 500 DEGREES K.
ELECTRICAL   VISCOSITY AND ELECTRICAL CONDUCTIVITY OF MOLTEN           0248
               GLASS.
ELECTRICAL   THE DIRECTIONAL DEPENDENCE OF ELECTRICAL                  0347
               CONDUCTIVITY IN METALS.
ELECTRICAL   PHOTOELECTRIC SENSITIVITY OF METALS AT LOW                0372
               TEMPERATURES AND ELECTRICAL PROPERTIES OF
               SPUTTERED FILMS.
ELECTRICAL   CRYSTAL GROWTH AND ELECTRICAL AND OPTICAL                 0421
               PROPERTIES OF GRAY TIN.
ELECTRICAL   ELECTRICAL CONDUCTION IN ZINC SULFIDE SINGLE              0515
               CRYSTALS.
ELECTRICAL   X-RAY, OPTICAL, AND ELECTRICAL PROPERTIES OF              0537
               BUILT - UP FILMS,
ELECTRICAL   CALCULATION OF THE RESONANT PROPERTIES OF                 0545
               ELECTRICAL CAVITIES.
ELECTRICAL   OPTICAL ABSORPTION PHOTOCONDUCTIVITY, ELECTRICAL          0554
               CONDUCTIVITY, AND HALL EFFECT IN GERMANIUM
               MONOSULFIDE.
ELECTRICAL   ON THE ELECTRICAL RESISTANCE OF MERCURY AT HIGH           0580
               TEMPERATURES AND HIGH PRESSURES, AND THE CRITICAL
               POINT OF MERCURY.
ELECTRICAL   REFLECTION AND REFRACTION OF ELECTRICAL WAVES BY          0625
               SCREENS OF RESONATORS AND BY GRIDS.
ELECTRICAL   ELECTRICAL PROPERTIES OF EVAPORATED CARBON FILMS.         0667
ELECTRICAL   THE ELECTRICAL PROPERTIES OF LEAD TELLURIDE FILMS.        0677
ELECTRICAL   THE INFLUENCE OF LIGHT ON THE ELECTRICAL                  0719
               RESISTANCE OF METALS.
ELECTRICAL   THE ELECTRICAL PROPERTIES OF TELLURIUM.                   0720
ELECTRICAL   AN INVESTIGATION OF CERTAIN ELECTRICAL PROPERTIES         0843
               OF OXIDE - COATED CATHODES.
ELECTRICAL   ION FORMATION AND DECAY IN A MERCURY RESONANCE            0876
               CELL AS EVIDENCED BY ELECTRICAL IMAGE FORCES.
ELECTRICAL   PART I.  DIELECTRIC LOSSES AT RADIO FREQUENCIES IN        0911
               LIQUID DIELECTRICS.  PART II.  THE ELECTRICAL
               PROPERTIES OF FLAMES CONTAINING SALT VAPORS FOR
               HIGH FREQUENCY ALTERNATING CURRENTS.  PART III.
               THE CONDUCTIVITY OF FLAMES FOR RAPIDLY ALTERNATING
               CURRENTS.
ELECTRICAL   THE THERMAL AND ELECTRICAL CONDUCTIVITIES OF              0933
               CARBON AND GRAPHITE AT LOW TEMPERATURES.
ELECTRICAL   THE ORIENTATION OF ELECTRICAL BREAKDOWN PATHS IN          1099
               SINGLE CRYSTALS.
ELECTRICAL   INVESTIGATIONS OF CERTAIN FREQUENCY DEPENDENT             1189
               ELECTRICAL PROPERTIES OF BIOLOGICAL MATERIALS.
ELECTRICAL   THE THERMAL AND ELECTRICAL CONDUCTIVITIES OF              1273
               LEAD - BISMUTH ALLOYS.
ELECTRICAL   THERMAL AND ELECTRICAL CONDUCTIVITIES OF TUNGSTEN         1456
               AND TANTALUM.
ELECTRICAL   ELECTRICAL AND OPTICAL PROPERTIES OF RUTILE               1498
               SINGLE CRYSTALS.
ELECTRICAL   THE ELECTRICAL CONDUCTIVITY OF AQUEOUS SOLUTIONS          1513
               OF STRONG ELECTROLYTES AT HIGH FREQUENCIES.
ELECTRICAL   THE EFFECT OF BOUNDARIES ON THE ELECTRICAL                1602
               PROPERTIES OF CERTAIN SEMICONDUCTORS.
ELECTRICAL   THE ORIENTATION OF ELECTRICAL BREAKDOWN PATHS IN          1626
               SINGLE CRYSTALS.
ELECTRICAL   THE CATHODO - CONDUCTIVITY OF ZINCBLENDE.  AN             1753
               EXPERIMENTAL INVESTIGATION OF THE EFFECT OF
               ELECTRON BOMBARDMENT ON THE ELECTRICAL
               CONDUCTIVITY OF ZINCBLENDE CRYSTALS.
ELECTRICAL   THE DECAY OF THE TRIPLET P LEVELS IN THE FIRST            1763
               EXCITED CONFIGURATION OF NEON DURING THE AFTERGLOW
               OF AN ELECTRICAL DISCHARGE.
ELECTRICAL   ELECTRICAL CONDUCTIVITY OF SINGLE CRYSTALS OF             1775
               BARIUM OXIDE AS A FUNCTION OF TEMPERATURE AND
               EXCESS BARIUM DENSITY.
ELECTRICAL   THE USE OF ELECTRICAL PULSE TECHNIQUES IN THE             1875
               STUDY OF THE MOBILITY OF GASEOUS IONS.
ELECTRICAL   THE ELECTRICAL PROPERTIES OF SEMICONDUCTORS               2019
               THROUGH THE SOLID - LIQUID TRANSITION.
ELECTRICAL   PULLING ELECTRONS OUT OF METALS BY INTENSE                2067
               ELECTRICAL FIELDS.
ELECTRICAL   THE ELECTRICAL RESISTIVITY OF COPPER ALLOYS AT LOW        2075
               TEMPERATURE.
ELECTRICAL   THE ELECTRICAL BEHAVIOR OF PLASTICALLY DEFORMED           2170
               (CONTINUED)
```

FIG. 5. Sample from an IBM Index: A Keyword Out of Context Index

the code first used led the editors to change to an identification code based on the journal citation rather than the author.[30] An example is the reference code "JIMT–0090–0172," referring to the *Journal of the Institute of Metals*, 90: 172.

The suggestion has been made that the code be constructed so that it could be used for shelving. "If the documents indexed were actually shelved and obtained by using Luhn's code or number, then the index would require only a single look-up."[40] Such indexes are preferred by users over the double look-up index, which requires the extra step of referring from the index to another listing for the information necessary to find the desired material. It has been objected that consistency in codes assigned on the basis of titles would be difficult to attain, but this difficulty might be overcome by a more arbitrary assignment of letters.

Special requirements of individual indexes have led to other code variations in order to increase the efficiency of the indexes. Experimentation with code variation will undoubtedly continue.

3. *Addition of classification information*

Regardless of the success of the permuted index, voices are still raised in favor of the classed system. For example, Guha, in a study of the arrangement of entries in a number of indexing periodicals, prefers a classified arrangement.[41] Campbell, speaking of keyword systems in general, agrees with Guha:

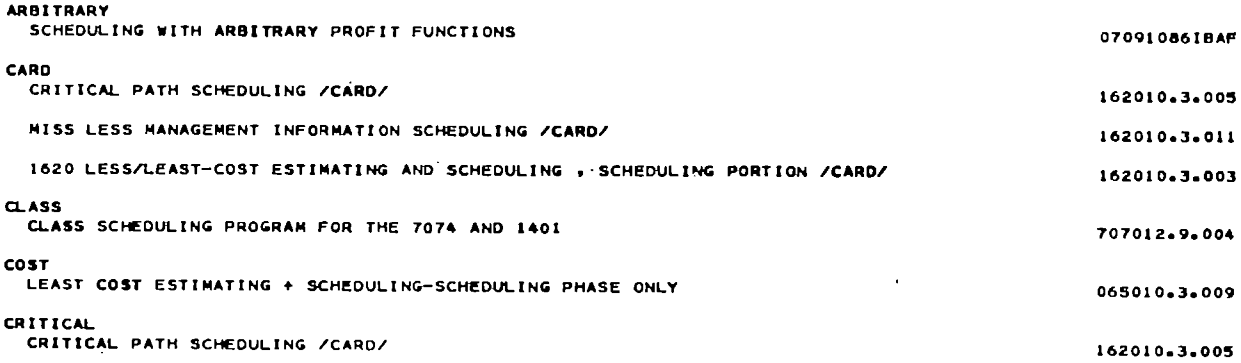
> For the user, the disadvantages of keyword systems are that he is confronted with a welter of words, which may or may not include all the words in which he puts his query, and that he needs, and rarely gets the "bird's eye view."[42]

Campbell proposes the use of "see" and "see also" cross references and also considers placing "the classification of keywords on to a hinged-panel strip index, close to the keyword index." This combination of keyword index and classified index "has most of the advantages of classification without many of its difficulties." For example, "because the classification does not now determine the location of any document in a file, or entry in an index, the need for mutually exclusive classes and of a single place for each concept vanishes."[42]

Kennedy also suggests that subject scatter, one of the characteristics, or shortcomings, of the permuted index, can be mitigated by cross referencing.[18] Examination of a KWIC index that used cross references indicated that the use was not as successful as it might have been because the type face of the cross references was the same as that of the entries, making it difficult for the user to determine which was which. However, this difficulty could probably be remedied by improvements of the type face.

4. *Combination of author index and title index: WADEX*

Another innovation made in the permutation index to achieve greater effectiveness has been the combination of



```
ARBITRARY
  SCHEDULING WITH ARBITRARY PROFIT FUNCTIONS                                          07091086IBAF

CARD
  CRITICAL PATH SCHEDULING /CARD/                                                     162010.3.005

  MISS LESS MANAGEMENT INFORMATION SCHEDULING /CARD/                                  162010.3.011

  1620 LESS/LEAST-COST ESTIMATING AND SCHEDULING , SCHEDULING PORTION /CARD/          162010.3.003

CLASS
  CLASS SCHEDULING PROGRAM FOR THE 7074 AND 1401                                      707012.9.004

COST
  LEAST COST ESTIMATING + SCHEDULING-SCHEDULING PHASE ONLY                            065010.3.009

CRITICAL
  CRITICAL PATH SCHEDULING /CARD/                                                     162010.3.005
```

FIG. 6. Sample from an Index Used at IBM: A Keyword Out of Context Index Using Full Title

subject and author entries in the WADEX index (word and author index). Since this index treats the authors' names as keywords, users need search only one index, not two — a convenience that has long been available in book indexes and library card files, but which seems to be new in the field of machine indexing. Also, users who remember papers by the names of the authors have another means of locating the indexed paper.[23]

5. *Improvement of type face*

The type face used in the KWIC indexes is an important aspect of the readability of the indexes. According to Balz, the small size of the type, after it has been reduced for printing, "Bothers everybody. . ." Balz reports that several things have been tried to improve the type. *Chemical Abstracts* has an upper and lower case chain in use, which may improve the readability. The chain will increase the cost of the index slightly.[43]

Use of bold face type also improves the readability of the indexes and would be of particular help in listing cross references, as mentioned above. In a machine printout, bold face "type" may be obtained by strikeovers.

• Improvements in the Preparation and Use of KWIC

The use of the KWIC concept and of KWIC indexes can be improved both by those who prepare the indexes and by those who use the prepared indexes. Those preparing the index can improve it by obtaining better titles and, perhaps, by the use of a thesaurus. Those using the prepared indexes can improve their use by a better understanding of the indexes.

1. *Improvement of titles*

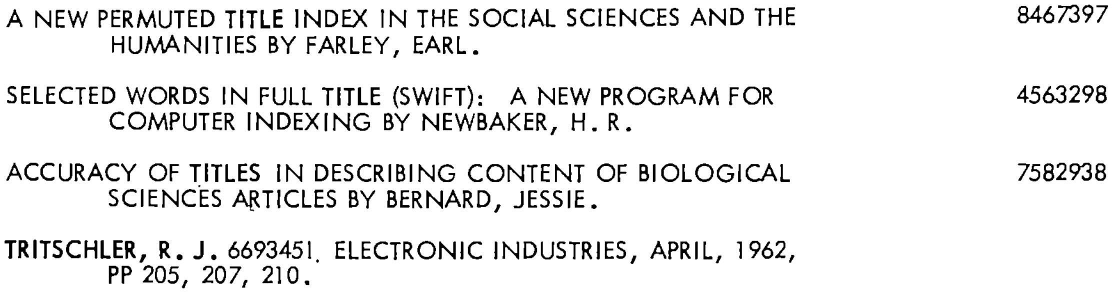
Most frequently mentioned in the literature of KWIC, perhaps, is the need for better titles. Studies have been made of the reliability of using titles as a basis for permutation indexing and of various other problems in titling.

Lane, in a survey of titles contained in ten periodical indexes, concludes that:

> in science and engineering the titles of articles usually describe or at least imply the contents of the articles. In non-technical fields titles reveal the contents less frequently; and in a general index such as *Readers' Guide* titles are indicative less than half the time.[33]

Sedano's statistical analysis of six indexes establishes a similar range from technical to general, with a similar range of title efficiency.[34] Title quality, or descriptiveness, correlates with the specificity of its literature.

Titles can serve two purposes: they can attract the attention of the reader or they can describe the subject of the article or publication. Also, when a secondary title is used with the primary title, the complete title can both



```
A NEW PERMUTED TITLE INDEX IN THE SOCIAL SCIENCES AND THE          8467397
        HUMANITIES BY FARLEY, EARL.

SELECTED WORDS IN FULL TITLE (SWIFT):  A NEW PROGRAM FOR           4563298
        COMPUTER INDEXING BY NEWBAKER, H. R.

ACCURACY OF TITLES IN DESCRIBING CONTENT OF BIOLOGICAL             7582938
        SCIENCES ARTICLES BY BERNARD, JESSIE.

TRITSCHLER, R. J. 6693451. ELECTRONIC INDUSTRIES, APRIL, 1962,
        PP 205, 207, 210.
```

FIG. 7. Sample of Improved Index

attract the reader and describe the subject. An example of a combined title, with primary and secondary titles, is:

The Shining Light:
A Blind Girl Meets Life and Finds Happiness
in San Francisco in Spite of Fire, Flood, Earthquake and
an Innocent Bystander

The phrase "The Shining Light" is the primary title and the phrase "A Blind Girl Meets Life and Finds Happiness in San Francisco in Spite of Fire, Flood, Earthquake and an Innocent Bystander" is the secondary title. The primary title is a catch title, the secondary title is a descriptive title with useful keywords.

Without too much difficulty, the convention of adding a descriptive secondary title containing keywords to a catch title could become a fixed stylization with the settled understanding that for permutation, only the secondary title would be indexed. By use of combined titles, the author would be able to supply catch titles to attract the reader and, at the same time, to supply a descriptive title with keywords for machine indexing.

Author participation in the writing of good titles is essential. The editors of *Biological Abstracts* approach the job of instructing authors, in a somewhat negative way, by supplying examples of very poor titles.

> *How should population surveys be made?* (Looks fine, only you might like to know before you look up this paper that it deals with fish.)
>
> *The problems of changing beliefs and attitudes.* (Would you guess this to be a general philosophic discussion? If so, you will discover instead some rather practical advice on the subject to leaders in wildlife programs.)
>
> *Be sure to jet to the skin.* (Aerospace biology? Guess again! Subject of the paper is control of blowflies on sheep.)[44]

Such titles could not be retrieved satisfactorily from a word-based index.

Kennedy, in suggestions to authors, recommends the following:

1. Consideration of the title as a one-sentence abstract
2. Use of specific terms
3. Provision of enough context to clarify the relationships between the selected keywords, but no more than enough
4. Balance of brevity and descriptive accuracy
5. Where possible, use of words rather than characters or notations that cannot be duplicated on standard key-punching and computing equipment
6. Filing of subjects in relation to titles to introduce general concepts into the word index.[28]

Herner approaches the problem of author participation from yet another and, ultimately, more critical direction. He indicates the degree of author participation already expected. As an example of implicit author participation, "in the specifications for papers for the 1963 ADI meeting there was the following requirement: 'The title must be composed with care and must contain at least six significant words.' "[26] As an example of explicit author participation, authors of papers presented at meetings of the Federation of American Societies for Experimental Biology were required to select indexing terms when preparing their titles. With such author participation in the machine-indexing process, Herner sees the dangers of standardization and conformity. These dangers, however, already exist in the "core" vocabularies of the different sciences and technologies. Whatever the ultimate objections to standardization and conformity may be, "The title must permutate or perish." [21]

The reliability of scientists in supplying titles was studied by Papier, using the criterion of retrievability. His sampling was small, consisting of five psychology articles, each sent to twenty scientists working in the general field of psychology. The scientists were asked to title the articles; then their titles were compared with the authors' titles. Comparison of words in the actual titles with words in titles supplied by the twenty scientists established a word frequency, which seems to be quite high: "53% of the scientists' words were found in the authors' titles, and 46% of the authors' words were found in the scientists' titles." [20] Use of a thesaurus raised the word frequency to 61% and 62%.

Another point to be considered in the preparation of titles is that it is possible for a title that seems to be good to be lost in the system and to be indexed inadequately or even not at all. An illustration of this is a fictional title by Brandenberg, "An Application Oriented Explanation of Machine Models Used by the Aerospace Companies." Made up entirely of stopwords from a particular computer-indexed information system, it would never enter the index at all—it would be stopped by the machine.[21]

2. *Use of a thesaurus*

Balz defines a thesaurus, as used in the field of information retrieval, as "a collection of authorized subject headings or 'descriptors.' These descriptors are arranged according to conceptual groups and fields, accompanied by an alphabetic index and containing pertinent scope notes and cross references." He explains further, "The idea of an official 'authority list' of subject headings or descriptors, however, is far from a new concept since 'authority lists' have been used by catalogers in libraries for many years." He goes on to define thesaurus in general library terms,

> The thesaurus, then, is the tool that catalogers or subject analysts use in describing the contents of documents in order that each may describe similar content consistently. It is also the device by which search requests may be worded to assure a relatively high degree of accuracy on retrieval.[45]

An example of a machine-made thesaurus may be seen in Fig. 8.

Papier is not convinced that supplying thesauri is of great value. In his study of indexing relevance, he says, "It can be said that providing thesauri in the sample

```
ACARICIDES
   (PEST CONTROL AND INHIBITING
   AGENTS)
   INCL: MITICIDES
   ALSO SEE: ANTIPEST IMPREGNANTS
      PARATHION
      PEST CONTROL

ACCELERATION
   (MECHANICS)
   ALSO SEE: DECELERATION

   ACCELERATION INTEGRATORS USE
      ACCELEROMETERS

ACCELERATION TOLERANCE
   (TOLERANCES)

ACCELERATORS
   (PARTICLE ACCELERATORS)
   ALSO SEE: BETATRONS
      CYCLOTRONS
      ELECTRON ACCELERATORS
      ELECTROSTATIC ACCELERATORS
      ION ACCELERATORS
      LINEAR ACCELERATORS
      PARTICLE ACCELERATORS
      PROTON ACCELERATORS
      SYNCHROTRONS
```

FIG. 8. Example of Machine-Made Thesaurus

studied could only increase the first try probability from 46% to a maximum of 62%."[20]

Storrer, like Papier, has a reserved attitude toward the need and value of a thesaurus with the KWIC index.

> The KWIC listing when used as an index minimizes the need of a thesaurus in seeking references on particular topics. Of course, a thesaurus is necessary to avoid problems arising from the use of synonyms, variance in spellings, and the ambiguities of identical words with altogether different meanings.[18]

Balz admits "that the use of a thesaurus in information retrieval work has not been universally agreed upon." Nonetheless, he "points out the need for a controlled list of descriptors and the pitfalls of operating an automated information retrieval system using uncontrolled natural language."

He states that,

> A system that does not use a thesaurus is based on the premise that words have precise meanings and that they do not derive any significance from context. However, some words do have more than one meaning, and thus there exist problems of interrelationships.[45]

The type of thesaurus proposed by Balz resembles a classed system. Specific ideas are grouped under broader levels so that narrow, related, subordinate concepts can be retrieved through broad areas.

> The purpose of a thesaurus is to enable retrievers of information to describe their information needs in terms used by the originators and indexers which is a basic requirement in the effective documentation of information.[45]

3. *Improvement of the use of KWIC*

For effective use of KWIC indexes and similar indexes, the user should be aware of the special characteristics of these indexes and should have an appreciation of the essential difference between these indexes and the subject-heading indexes.

The user should realize that he will need to look for synonyms of the words that most precisely describe the subject for which he is searching. For example, when searching for literature on the KWIC index, reference must be made to the term "keyword-in-context" as well as to "KWIC."

The user should realize that he may need to look from the specific to the general when using KWIC indexes rather than from the general to the specific, as is the case when using a subject-heading index. For example, literature on KWIC is indexed under the general terms "permutation index," "permuted indexing," and "indexing."

Also, the user must be aware of terms used in relationship to the subject for which he is searching, even though those terms are not in hierarchal relationship to each other. The ease with which a narrow subject may be located in the KWIC indexes should not detract the user from other possible leads to relevant bits of information. For example, reference should be made to the related term "titles" in a search for literature on the KWIC index.

Librarians, scientists, technologists, and other professional people using KWIC indexes should all be aware of the special characteristics of the indexes. It is particularly important for the librarian using KWIC indexes to have a greater awareness of special vocabularies than he needs when using a subject heading index.

● Usage of the KWIC Index

Perhaps no aspect of the KWIC index is as controversial as its usage. The KWIC index was created to "cope with the problems of timeliness."[30] It was conceived of as a current awareness tool and it is primarily used as such. It has been spoken of as "the best replacement for the old library bulletin, for getting things out to people quickly."[46] Luhn spoke of the KWIC system and its usage as follows:

> (1) The principal merit of the method is timeliness. The KWIC system lends itself to index production in the shortest possible time with a minimum of effort.
> (2) The proper objective of KWIC indexes is to increase among their readers an awareness of current research.
> (3) The usefulness of these indexes is of a temporary nature. Ideally, they should be superseded . . . by "an instrument prepared with care in due course, incorporating all those features which will enhance its usefullness as a permanent tool of reference."[47]

The semimonthly index, *Chemical Titles,* represents:

> a faithful adaptation of Luhn's proposal; it is a disseminating index rather than a retrieval index. Aimed

> at increasing 'current awareness,' each semimonthly *Chemical Titles* is a concordance to articles selected from some 600 periodicals. It is not an index to *Chemical Abstracts*. Papers are likely to be listed in *Chemical Titles* even before their abstracts are published. It is apparent that *Chemical Titles* is not intended to take the place of thorough indexing-in-depth.[47]

Librarymaster Services uses KWIC for new items announcement bulletins because it is "the most economical listing . . . it is in effect a hybrid subject and title catalog."[48] A similar permuted title system used at Bell Telephone Laboratories provides "a current awareness bulletin with a highly useful subject structure."[49]

Luhn, a few months after he proposed the KWIC index, also proposed an accompanying service system which he named Selective Dissemination of Information (SDI). He explained the need for such a system in this manner:

> Effective dissemination of scientific information has of late become the subject of major interest and concern because of the realization that it is apt to play a decisive part in the race for leadership in technological accomplishments, be it among nations or be it among organizations and businesses within a nation. It is felt that if discoveries and new developments can promptly and exhaustively be brought to the attention of scientists and engineers at large, technological progress may be accelerated. This feeling is obviously born from a conclusion that presently existing means of scientific communication are inadequate, a condition which has invited the attention of government and institutional leaders and bodies.[50]

Luhn's method of selective dissemination of information consists of a machine-comparison of a pattern of keywords characterizing a new document with an interest profile of a user. If there is enough similarity between the two, the user is notified by card. If the user wishes to have the document, he returns a stub from the card and the document is sent to him. SDI is the subject of research and experimentation by Chemical Abstracts Service[51] and is in routine use at IBM.

SDI can be considered to be an extension of the same principle that underlies the KWIC index; that is, the rapid dissemination of current information from source to interested user.

KWIC indexes are used as retrospective searching tools as well as current awareness tools. As stated by Balz and Stanwood of IBM, "KWIC indexing allows rapid preparation of current awareness bulletins and at the same time provides a retrospective multi-aspect search facility."[51] A similar opinion is offered by Kennedy in a description of Bell Telephone Laboratories use of the permuted index:

> Again by tape merge or punched deck re-runs, the production of cumulated and up-dated index volumes on a continuous or periodic basis, say semiannually, is simple and cheap. The original investment in the announcement bulletin is thereby exploited to provide a multi-aspect retrospective search facility.[49]

The *Biological Abstracts* adaptation of the KWIC index, which is called BASIC (acronym for Biological Abstracts Subjects in Context), is designed "to meet not only a need for 'current awareness,' but also the need for a permanent reference tool."[47] Enthusiasm for its use by scientists may be represented by the statement of one who wrote, "This is as great an advance to biology as the electron microscope."[47]

WADEX (acronym for Word and Authors Index), which is an index to *Applied Mechanics Reviews*, "is intended for retrospective search rather than for current awareness information, though it could also be used for the latter purpose."[23]

It is interesting and significant to note that, as the use of the KWIC index deviates farther and farther from a simple current awareness bulletin, it becomes more necessary to change or adapt it to the enlarged usage of retrospective search. Almost always, the first change to be made is in the practice of editing, to make KWIC more acceptable as a permanent book index. Balz and Stanwood make this statement about KWIC:

> This system produces indexes more rapidly, accurately, and with greater definition of subject content than other forms of indexing; especially when provision is made to add descriptive words to titles that do not in themselves convey the subject content of the papers they represent.[51]

Kennedy, of Bell Telephone Laboratories, writes:

> There are . . . no rules in the game which say that human contributions to mechanized indexing are illegal! that data to be keypunched for indexing must come from the title alone. Several steps for adding to the coverage or convenience of a permuted index can be taken without compromising the essential merits of mechanization. At the editorial scan stage titles which appear on the basis of this quick inspection to be weak or unclear might be supplemented, say by marking a word or words in the document abstract for keypunching . . .[13]

Biological Abstracts uses "vocabulary management" and "supplementation" of titles in its index which is used for current awareness and for cumulation.[29] WADEX, which departs from the classic KWIC format in a number of ways, also uses editing in the manner described as follows:

> As the first step, the titles as they appear in the magazine are edited by an engineer or scientist to remove or change all those features which cannot be properly handled by the keypunch and printout equipment. . . . These titles are then keypunched. After verification, the cards are fed into the IBM 1401 (Program A) which prints out their contents for another human post-editing. After corrections from this editing are inserted, machine processing begins.[23]

On the use of editing, Herner takes the viewpoint that:

> In all probability, unless the existing technology changes drastically, any permuted index that resembles a conventional book index will be the product of either

> human operators alone or of machines working in tandem with human operators, whose job it would be to make sure that the machines make sense.[52]

• The Future of KWIC Indexes and of the KWIC Concept

The KWIC index may some day find a place in a national index that might be part of a national information center patterned after the center in Russia, or it may find a place in regional information centers or in discipline-oriented centers. According to Veyette, such centers should concern themselves with current literature, and should have

> an active program of rapid, automatic, selected dissemination. A system to accomplish such a program could be the Selective Dissemination of Information system. . . . Such a system would notify scientists, engineers, management personnel, educators, governmental employees, and the private citizen, if desired, of reports, articles, and other papers of interest to him.[5]

KWIC indexes, with their emphasis on currency and by their link with SDI, will fit easily into such an information system.

The future for KWIC includes the plans to replace card-punch operation with optical scanners, making the preparation of the index even faster than it is today. Also, plans for using the "Echo" satellites to link information centers around the world, in a worldwide drive toward immediacy in information dispersion, will surely provide a place for KWIC indexes and for the KWIC concept.

References

1. Luhn, H. P. 1959. *Keyword-in-Context Index for Technical Literature (KWIC Index)*. RC–127. IBM Corporation, Yorktown Heights, N. Y. (Aug.) Also, *Am. Doc.,* 9: 288–295. (Oct.) 1960.
2. Citron, J., Hart, L., and Ohlman, H. 1958. *A Permutation Index to the Preprints of the International Conference on Scientific Information*. SP 44. System Development Corporation, Santa Monica, Calif. (Nov.) Also a revised edition, December 15, 1959.
3. *New Scientific Journal Uses Machine Indexing*. 1960. Current Comments, *Law Library Journal,* 53: 223–224.
4. Turner, L., and Kennedy, J. H. 1961. *System of Automatic Processing and Indexing of Reports*. UCRL–6510. Lawrence Radiation Laboratory, Livermore, Calif. (July).
5. Veyette, Jr., J. H. Information Retrieval. *The American Behavioral Scientist,* V: 15–20.
6. Resnick, A. 1961. Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents. *Science,* 134: 1004–1006. (Oct. 6.)
7. *General Information Manual; Keyword-in-Context (KWIC) Indexing*. 1962. Technical Publications Department, IBM Corporation, White Plains, N. Y. E 20–8091.
8. Balz, C. F. and Stanwood, R. H. 1962. *On Preparing Information for KWIC Indexing (IBM 7090)*. Federal Systems Division, Space Guidance Center, IBM Corporation, Owego, N. Y. 62–816–729. (Jan. 15.)
9. Giallanza, F. V., and Kennedy, J. H. 1962. *Key-Word-in-Title (KWIT)* Index for Reports. UCRL-6782. Lawrence Radiation Laboratory, Livermore, Calif. (May 14.)
10. Swartz, M. D., and Farley, E. 1962. (Untitled as research in progress at University of Oklahoma, Norman, Oklahoma.) Reported in *Current Research and Development in Scientific Documentation,* 11: 171. National Science Foundation. (Nov.)
11. Lazarow, A., Newill, V., Izzo, J., et. al. 1962. (Research in progress for American Diabetes Association, New York.) Reported in *Current Research and Development in Scientific Documentation,* 11: 63. National Science Foundation. (Nov.)
12. Stanwood, R. H. 1962. *The Merge System of Information Dissemination, Retrieval and Indexing Using the IBM 7090 DPS.* 62–825–441. Space Guidance Center, IBM Corporation, Owego, N. Y. (Sept.)
13. Kennedy, R. A. 1962. Library Applications of Permutation Indexing. *J. Chem. Doc.,* 2: 181–185. (July)
14. Kraft, D. H. 1964: *A Comparison of Keyword-in-Context (KWIC) Indexing of Titles with a Subject Heading Classification System.* Paper read at the Annual Convention of the American Documentation Institute, Dec. 13, 1962. *Am. Doc.,* 15: 48–52. (Jan.)
15. Kurmey, W. J. 1962. (Untitled research project in progress at University of Chicago, Chicago, Ill.) Reported in *Current Research and Development in Scientific Documentation,* 11: 169–170. National Science Foundation. (Nov.)
16. Black, J. D. 1962. The Keyword: Its Use in Abstracting, Indexing and Retrieving Information. *Aslib Proc.,* 14: 313–321. (Oct.)
17. Veilleux, M. P. 1962. *Permuted Title Word Indexing: Procedures for Man/Machine System.* (Paper presented at the Third Institute on Information Storage and Retrieval, Feb. 14, 1961.) *Machine Indexing: Progress and Problems,* American University, Washington, D. C., pp. 77–111.
18. Storrer, R. L. 1963. *KWIC and Dirty Information Retrieval.* Paper read at the 1620 Users Group, Eastern and Midwestern Joint Meeting, Pittsburgh, Pa. (Oct. 14–16.)
19. Ruhl, M. J. 1963. *Chemical Documents and Their Titles: Human Concept Indexing vs KWIC-Machine Indexing.* Paper read at the 144th National Meeting, American Chemical Society, Los Angeles, Calif. (Apr. 2.)
20. Papier, L. 1963. *Reliability of Scientists in Supplying Titles; Implications for Permutation Indexing. Aslib Proc.,* 15: 333–335. (Nov.)
21. Brandenberg, W. 1963. Write Titles for Machine Index Information Retrieval Systems. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 57–58.
22. Bernard, J., and Shilling, C. W. 1963. Accuracy of Titles in Describing Content of Biological Sciences Articles. *Biol. Abstr.,* BSCP Communique 10–63. Philadelphia, Pa. (May).

23. Ripperger, E. A., Wooster, H., Juhasz, S., and Roach, F. 1963. WADEX (Word & Authors Index); A New Tool in Literature Retrieving. *Appl. Mech. Rev.*, pp. 951–956. (Dec.)
24. Newbaker, H. R., and Savage, T. R. 1963. Selected Words in Full Title (SWIFT): A New Program for Computer Indexing. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 87–88.
25. Chonex, N., Chonex, A., and Iung, J. 1963. Physindex: An Auto-Indexed Current List of Physics Literature Produced on IBM 1401 Computer. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 31–32.
26. Herner, S. 1963. Effect of Automated Information Retrieval Systems on Authors. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 101–102.
27. Conolly, T. F. 1963. Author Participation in Indexing — From Primary Publication to Information Center. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 35–36.
28. Kennedy, R. A. 1963. Writing Informative Titles for Technical Papers — A Guide to Authors. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 133–134.
29. Parkins, P. V. 1963. Approaches to Vocabulary Management in Permuted-Title Indexing of Biological Abstracts. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 27–28.
30. Freeman, R. R., and Dyson, G. M. 1963. Development and Production of *Chemical Titles,* a Current Awareness Index Publication Prepared with the Aid of a Computer. *J. Chem. Doc.*, 3: 16–20. (Jan.)
31. Freeman, R. R. 1963. Automatic Retrieval and Selective Dissemination of References from Chemical Titles: Improving the Selection Process. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 213–214.
32. East, H., Shaw, T. N., and Smith, A. C. 1963. Letter to the Editor; Keyword in Context Indexes. *Aslib Proc.*, 15: 31–32. (Jan.)
33. Lane, B. B. 1964. Key Words in — and out of — Context. *Spec. Lib.*, 55: 45–46. (Jan.)
34. Sedano, J. M. 1964. *Keyword-in-Context (KWIC) Indexing: Background, Statistical Evaluation, Pros and Cons, and Applications.* Unpublished thesis, University of Pittsburgh.
35. Fischer, M. F. 1964. *History and Use of the KWIC Index Concept,* a thesis. San Jose State College, San Jose, Calif.
36. Stevens, M. E. 1965. *Automatic Indexing: A State-of-the-Art Report,* NBS Monograph 91. GPO. U. S. Department of Commerce, National Bureau of Standards, Washington, D. C. (Mar. 30.)
37. Farley, E. 1963. A New Permuted Title Index in the Social Sciences and the Humanities. *Spec. Lib.*, 54: 557–562. (Nov.)
38. Metcalf, J. W. 1965. *Alphabetical Subject Indication of Information* (Rutger's Series on Systems for the Intellectual Organization of Information, Vol. 3), Rutger's, Graduate School of Library Service.
39. Janaske, P. C., editor. 1962. *Information Handling and Science Information: A Selected Bibliography 1957–61.* American Institute of Biological Sciences, B.S.C.P., Washington, D. C.
40. Youden, W. W. 1963. Characteristics of Programs for KWIC and Other Computer-Produced Indexes. *Automation and Scientific Communication,* American Documentation Institute, Washington, D. C., pp. 331–332.
41. Guha, B. 1963. Arrangement of Entries in Four Indexing Periodicals. *Documentation Periodicals,* pp. 119–126.
42. Campbell, D. J. 1963. Making Your Own Indexing System in Science and Technology; Classification and Keyword Systems. *Aslib Proc.*, 15: 282–303. (Oct.)
43. Personal correspondence, Charles F. Balz to the author, April 14, 1964.
44. For Want of a Title — A Paper was Lost. 1962. *Biol. Abstr.*, 37: xii. (Jan. 15.)
45. Balz, C. F. 1962. *The Need for a Thesaurus in Automated Information Retrieval.* 62–825–481. Federal Systems Division, Space Guidance Center, IBM Corporation, Owego, N. Y. (Sept.)
46. Personal correspondence, Dr. I. A. Warheit to the author, February 26, 1964.
47. Lewis, R. F. 1964. "KWIC . . . Is It Quick?" *Bulletin of the Medical Library Association,* 52: 142–147. (Jan.)
48. *Librarymaster Services Manual,* Librarymaster Services, Oakland, Calif., n.d.
49. Kennedy, R. A. 1963. Indexing by Machine. *Mater. Res. Std.*, 3: 752–753. (Sept.)
50. Luhn, H. P. 1959. *Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment.* 17.010. ASDD, IBM Corporation, Yorktown Heights, N. Y. (Nov. 30.)
51. Balz, C. F., and Stanwood, R. H. 1962. *Some Applications of the KWIC Indexing System.* 62–825–475. Federal Systems Division, Space Guidance Center, IBM Corporation, Owego, N. Y. (June 15.)
52. Herner, S. 1963. *Deep Subject Indexing by Manual Permutation Methods,* Air Force Office of Scientific Research, Washington, D. C. (Oct. 22.)

Characteristics and Use of Personal Indexes Maintained by Scientists and Engineers in One University[1]

Interviews with 75 graduate school faculty members in the science and engineering departments at Florida State University have revealed that 46 of the interviewed faculty members maintain personal indexes. The structure of these personal indexes, their size, rate of growth, frequency of use, physical form, and other characteristics are given and discussed.

G. JAHODA, RONALD D. HUTCHINS, and ROBERT R. GALFORD

Library School
Florida State University
Tallahassee, Florida

Personal indexes are organized collections of documents and/or homemade references to documents that the researcher keeps in his office. Information gathering habit studies have shown that a significant portion of researchers maintain personal indexes. Studies by Fishenden, Tornudd and Hogg and Smith, for example, have brought out the fact that 45% (1), 57% (2) and 66% (3), respectively, of surveyed scientists had and/or used personal indexes. Zwemer has found that nearly every scientist surveyed in a recent study kept a personal file in the way of reprints, abstracts, or notes on cards, and that the average rate of growth of 26 such collections is 330 items per year (4). In another recent study of the information needs of Department of Defense scientists and engineers, 17% of the interviewed scientists and engineers used personal files as their first source of information, while 51% of the interviewed scientists and engineers relied on their local environment — personal files, departmental files, and colleagues — as a first source of information (5). Heller and Wallace suggested that information specialists be used to assist in the preparation of personal indexes. They describe the preparation of personal indexes, printed with the aid of computers, for professional and administrative personnel in one organization, the Systems Development Corporation (6, 7). A study now being done at Florida State University carries this suggestion one step further. Personal indexes will be prepared for a group of researchers in science and engineering and the use of these indexes will be studied. The study is being carried out in several stages. The first stage, the one that has been completed and is described in this report, consisted of a survey of personal indexes now in use.

Information about the personal indexes maintained and used by Florida State University graduate faculty members in science and engineering was obtained by means of personal interviews. Faculty members to be interviewed were selected from the 1964–65 Graduate Bulletin of Florida State University. The Chemistry, Food and Nutrition, Biology, Geology, Mathematics, Physics, Meteorology, and Statistics Departments, and the School of Engineering were included in this study. Only faculty members in the Graduate School were selected since it was believed that these members were most active in research. Heads of departments (with one exception) were excluded because their activities were considered to be predominantly administrative rather than research oriented. The original sample of 105 researchers that was selected in September 1964 was increased by seven (new appointments to the faculty) and reduced by 37 (resignations, leaves of absence, or unwillingness to participate in the study). Thus 75 researchers have been interviewed, with 46 researchers having a personal index and 29 researchers not having a personal index. The interviews were conducted from September 1964 to July 1965 and lasted an average of 55 minutes with researchers who maintain a personal index. No record of time was kept for interviews of researchers who did not have a personal index. The results of the interviews are given in Tables 1–16.

[1] Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant Number 895–65.

Table 1. Distribution by Departments.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
With Personal Index	16 = 84%	10 = 67%	5 = 71%	4 = 33%	4 = 50%	7 = 50%	46 = 66%
Without Personal Index	3 = 16%	5 = 33%	2 = 29%	8 = 67%	4 = 50%	7 = 50%	29 = 34%

Table 2. Distribution by Academic Rank.

	Professor	Associate Professor	Assistant Professor	Total
With Personal Index	20 = 71%	19 = 63%	7 = 41%	46 = 66%
Without Personal Index	8 = 29%	11 = 37%	10 = 59%	29 = 34%

Table 3. Number of Documents in Personal Index.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Less than 1,000	6 = 38%	2 = 25%		3 = 100%	1 = 33%	4 = 67%	16 = 40%
1,000 – 2,000	1 = 6%	5 = 63%	1 = 25%		1 = 33%		8 = 20%
2,000 – 3,000	4 = 25%	1 = 12%	1 = 25%		1 = 33%		7 = 18%
3,000 – 4,000	1 = 6%		1 = 25%				2 = 5%
4,000 – 5,000	1 = 6%					2 = 33%	3 = 8%
5,000 – 10,000	3 = 19%		1 = 25%				4 = 10%
Number of respondents = 40.							

Table 4. Rate of Growth Per Month.*

	Biology	Chemistry	Geology	Math	Physics	Other	Total
1 – 5	2 = 14%			2 = 67%			4 = 12%
6 – 15	8 = 57%	3 = 50%		1 = 33%		1 = 20%	13 = 38%
16 – 30	3 = 22%	2 = 33%	2 = 50%		2 = 100%	2 = 40%	11 = 32%
31 – 50		1 = 17%				1 = 20%	2 = 6%
51 – 75						1 = 20%	1 = 3%
76 – 100	1 = 7%		2 = 50%				3 = 9%
Number of respondents = 34.							

* In number of documents.

Table 5. Age of Index in Number of Years.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Less than 1	1 = 6%						1 = 2%
1 – 5	1 = 6%	2 = 22%	2 = 40%	2 = 50%	2 = 50%	3 = 43%	12 = 27%
6 – 10	2 = 13%	3 = 33%	1 = 20%	1 = 25%	2 = 50%	3 = 43%	12 = 27%
11 – 15	5 = 31%	3 = 33%	1 = 20%	1 = 25%			10 = 22%
16 – 20	3 = 19%		1 = 20%			1 = 14%	5 = 11%
More than 20	4 = 25%	1 = 11%					5 = 11%
Number of respondents = 45.							

Table 6. Frequency of Updating.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Daily	1 = 7%	1 = 10%	2 = 50%			1 = 33%	5 = 13%
Weekly	1 = 7%		1 = 25%		1 = 33%		3 = 8%
Monthly	4 = 27%	1 = 10%	1 = 25%				6 = 16%
As collected	9 = 60%	7 = 70%		3 = 100%	2 = 67%	2 = 67%	23 = 60%
2–3 times per year		1 = 10%					1 = 3%
Number of respondents = 38.							

Table 7. Types of Documents Included.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Journal articles	16 = 100%	10 = 100%	5 = 100%	4 = 100%	4 = 100%	7 = 100%	46 = 100%
Conference papers	11 = 69%	5 = 50%	4 = 80%	2 = 50%	3 = 75%	3 = 43%	28 = 61%
Government reports	5 = 31%	5 = 50%	5 = 100%	1 = 25%	2 = 50%	5 = 71%	23 = 50%
Patents		3 = 30%				1 = 14%	4 = 9%
Technical correspondence	1 = 6%	3 = 30%	1 = 20%	1 = 25%			6 = 13%
Lecture notes			1 = 20%		1 = 25%		2 = 5%
Trade literature			1 = 20%		1 = 25%		2 = 5%
Books in their index	2 = 13%	1 = 10%		1 = 25%		1 = 14%	5 = 11%
Thesis-dissertation	1 = 6%						1 = 2%
Seminar notes		1 = 10%					1 = 2%
Number of respondents = 46.							

Table 8. Physical Arrangement of Original Documents.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Subject *	8 = 50%	9 = 90%	4 = 80%	3 = 75%	1 = 25%	1 = 16%	26 = 58%
Type of document	9 = 56%	5 = 50%	2 = 40%	1 = 25%	1 = 25%	3 = 50%	21 = 47%
Author	8 = 50%	3 = 30%		1 = 25%	1 = 25%	2 = 33%	15 = 33%
Date	1 = 6%		1 = 20%				2 = 5%
Accession number	5 = 31%				1 = 25%		6 = 13%
Number of respondents = 45.							

* Includes several which are arranged by subject with subarrangement by author.

Table 9. Access Points.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Subject	14 = 88%	10 = 100%	4 = 80%	3 = 75%	2 = 50%	4 = 57%	37 = 80%
Author	11 = 69%	4 = 40%	2 = 40%	3 = 75%	4 = 100%	4 = 57%	28 = 61%
Title	2 = 13%						2 = 5%
Project	4 = 25%	1 = 10%	1 = 20%	1 = 25%		1 = 14%	8 = 17%
Number of respondents = 46.							

Table 10. Average Number of Access Points.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
1	8 = 50%	6 = 60%	5 = 100%	4 = 100%	2 = 50%	4 = 57%	29 = 63%
2	3 = 19%	2 = 20%			1 = 25%	1 = 14%	7 = 15%
3	2 = 13%					1 = 14%	3 = 7%
4	1 = 6%				1 = 25%		2 = 5%
5	2 = 13%	2 = 20%					4 = 9%
6 - 8						1 = 14%	1 = 2%
Number of respondents = 46.							

Table 11. Types of Subject Indexes.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Alphabetical subject	4 = 31%	1 = 13%	2 = 100%		2 = 100%	1 = 33%	10 = 24%
Coordinate	3 = 23%	2 = 25%				2 = 67%	7 = 17%
Broad subject or classified	10 = 76%	7 = 88%	2 = 100%	3 = 100%	1 = 50%	1 = 33%	24 = 59%
Number of respondents = 31.							

TABLE 12. Physical Form of Index.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
On cards	12 = 75%	6 = 67%	4 = 80%	1 = 25%	3 = 75%	4 = 100%	30 = 70%
File folders	7 = 44%	5 = 56%	2 = 40%	3 = 75%	2 = 50%	2 = 50%	21 = 49%
Page form	1 = 6%						1 = 2%
Pamphlets, boxes	2 = 13%						2 = 5%
Number of respondents = 42.							

TABLE 13. Amount of Bibliographic Information Included in Index Entries on Cards.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Citation or accession No.	8 = 67%	3 = 50%	2 = 50%	1 = 100%	1 = 33%	1 = 25%	16 = 53%
Citation and keywords	1 = 8%				2 = 67%		3 = 10%
Citation and abstract	3 = 25%	3 = 50%	2 = 50%			3 = 75%	11 = 37%
Number of respondents = 30.							

TABLE 14. Frequency of Use.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Daily	8 = 50%	6 = 67%	2 = 50%	2 = 50%	1 = 25%	2 = 33%	21 = 49%
Twice weekly	5 = 31%	3 = 33%				1 = 17%	9 = 21%
Weekly			1 = 25%	2 = 50%	2 = 50%	2 = 33%	7 = 16%
Twice monthly			1 = 25%				1 = 2%
Monthly						1 = 17%	1 = 2%
Sporadic	3 = 19%				1 = 25%		4 = 9%
Number of respondents = 43.							

TABLE 15. Age Group of Most Active Material in Number of Years.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Up to 2	3 = 19%	1 = 10%		1 = 25%	1 = 25%		6 = 13%
Up to 5	6 = 38%	6 = 60%	2 = 40%			4 = 57%	18 = 39%
Up to 10	2 = 13%	3 = 30%		1 = 25%	2 = 50%	2 = 29%	10 = 22%
Older than 10	2 = 13%		3 = 60%	1 = 25%	1 = 25%	1 = 14%	8 = 17%
All equal	3 = 19%			1 = 25%			4 = 9%
Number of respondents = 46.							

TABLE 16. Shortcomings and Desired Improvements Suggested by Researchers.

	Biology	Chemistry	Geology	Math	Physics	Other	Total
Too time consuming to prepare	5 = 56%	1 = 33%			1 = 50%	1 = 50%	8 = 42%
Inconsistencies in indexing	2 = 22%	2 = 67%	2 = 67%				6 = 32%
Not enough access points	2 = 22%	1 = 33%					3 = 16%
Lacks subject approach	1 = 11%				1 = 50%	1 = 50%	3 = 16%
Collection inadequate			1 = 33%				1 = 5%
Not up to date	1 = 11%		1 = 33%				2 = 11%
Subject headings too detailed	1 = 11%						1 = 5%
Alphabetical arrangement unwieldy	1 = 11%						1 = 5%
Number of respondents = 19.							

The first two tables give the distribution of researchers with and without a personal index by department and rank. The subsequent tables characterize the personal indexes with the tables being arranged in the order of the questions in the interview schedule. For each table the number and the percentage of responses are given by individual departments with the exception of the Departments of Food and Nutrition, Meteorology, and the School of Engineering. The small number of interviewed researchers who had personal indexes in these areas led us to group their answers under "other." Several questions yielded more than one answer per researcher, e.g., types of documents included, and it is for this reason that the percentage figures for these questions add up to more than 100.

The surveyed personal indexes do not constitute a statistically representative sample because researchers in only one university have been studied and because the sample is too heterogeneous in terms of subject interests. Nevertheless, a number of observations will be made about the personal indexes, observations that might be tested on a larger and more homogeneous sample. Most collections are relatively small in size and are not now growing at a very rapid rate. Of the 66% of the interviewed researchers who had a personal index, 60% contained 2,000 or fewer documents, 78% contained 3,000 or fewer documents. Eighty-two per cent of the indexes grew at the rate of 30 or fewer documents per month. This raises the question of whether there are relatively few basic documents of interest to the researcher or whether the library is able to fill the researcher's need for basic documents, to consider only two possibilities. The researchers' personal indexes are frequently updated, and this appears to be an indication of the importance of this tool to the researcher. Eighty-one per cent of the indexes are updated at least weekly (if we assume that updating "as collected" means at least weekly).

While all indexes in the sample included journal articles (including reprints and preprints), other forms of publications were not included in a number of files. This was not surprising in the case of patents since the researchers' interests were concentrated on the basic rather than applied sciences. The small number of indexes that included technical correspondence (13%) appears to be an indication that scientific information of more than ephemeral value is not recorded in this form. Only 5% of the indexes included trade literature (mostly product, equipment, or chemical catalogs), but a subsequent check of the offices of 11 researchers showed that 10 out of 11 researchers kept trade literature in their offices (though not in an organized form). Only 2% of the researchers indicated during the interviews that they include theses or dissertations in their personal indexes. However, subsequent visits to 11 offices indicated that seven out of 11 researchers had theses and dissertations on their shelves. The response on the age of the most active material is also worth noting. Thirty-nine per cent of the researchers considered material up to five years old most active. Material up to two years old was considered most active by 13% of the researchers. Of the 80% of the researchers who had a subject approach to their indexes, 59% used a broad subject (classified) arrangement, 24% an alphabetic subject index, and 17% a coordinate index.

Shortcomings in or problems with the personal index were listed by 19 researchers. Forty-two per cent who responded to this question considered the time devoted to prepare their index as excessive, 32% complained of inconsistencies in indexing (a problem of professional indexers as well), 16% desired more access points, 16% missed a subject approach (this represents three out of 19 researchers who answered this question and did not have a subject approach), and only 5% considered their collection inadequate. Almost half (49%) of the researchers used their indexes daily, another 39% used their personal indexes at least weekly, according to the interviews.

Researchers who indicated daily use of their personal indexes were asked to participate in the next stages of the study. This consists of collecting case histories of personal index use, analyzing these individual case histories to determine what type or types of indexes appear to be most suitable, designing personal indexes based on this analysis, and studying the use of these indexes. The collection and analysis of case histories of use of seven personal indexes are now underway.

References

1. Fishenden, R. M. 1959. Methods by which research workers find information. 1: 169. *In* Proc. of International Conference on Scientific Information. National Academy of Sciences, National Research Council, Washington, D. C.
2. Tornudd, E. 1959. Study of the use of scientific literature. 1: 55. *In* Proc. of the International Conference on Scientific Information. National Academy of Sciences, National Research Council, Washington, D. C.
3. Hogg, I. H., and Smith, J. R. 1959. Information and literature use in a research and development organization, p. 140. *In* Proc. of the International Conference on Scientific Information. National Academy of Sciences, National Research Council, Washington, D. C.
4. Zwemer, R. L. 1963. A biological information survey: Discussion and observation. *In* Studies in Biological Literature and Communications, No. 2. Biological Abstracts, Inc., Philadelphia.
5. Auerbach Corporation. 1965. DOD user needs study; Phase I., Vol. I, Section 1, p. 12. Dept. of Defense, Advanced Research Projects Agency, Final Technical Report 1151–TR–3. (Also AD–615 501.)
6. Heller, E. W. 1963. Applied information management system, pp. 161–162.. *In* Annual Meeting of the American Documentation Institute, Chicago, October, 1963; Short Papers, Part 2. American Documentation Institute, Washington, D. C.
7. Wallace, E. M. 1964. Experience with EDP support of individuals' file maintenance. *In* Proc. of the Am. Doc. Institute, Parameters of Info. Science, 1: 259–261. American Documentation Institute, Washington, D. C.

State of the Art of Computers in Commercial Publishing

Despite excessive glamorizing of the role of computers in publishing, truly economical applications are beginning to emerge. Many types of directories, indexes, cumulated bibliographies, and cumulated library book catalogs can today be put through a computer in less time and at less cost than for conventional typesetting. Examples of successful applications are given, after taking up one by one the many system design decisions that must be made in the areas of input hardware, computer hardware, computer software, and output hardware. Major attention is given to the proofreading problem, because line printers are still far from ideal for producing the equivalent of galley proofs.

JOHN MARKUS[1]

McGraw-Hill, Inc.
New York, New York

● Introduction

This paper starts with a dream. In this dream there is a huge room, filled with hundreds of blondes, brunettes, and redheads. They are all retyping manuscripts in rhythm, to the tune of the Stars and Stripes Forever, on tape-punching typewriter keyboards. Nobody worries about justification, punctuation, widows, rivers, or the other little details of typography that plague a Linotype operator. The typewriter hard copy is proofread, correction tapes are punched, and all the tapes are fed into a computer at 500 or 1,000 characters per second.

The computer justifies and hyphenates the copy as per typographical specifications. It also inserts repetitive words, supplies headings, makes up pages, eliminates bad breaks, calculates total length, and even spreads out or squeezes together the lines to eliminate blank pages in the last printing form.

The final output tape of the computer is error-free, because our computer does not make mistakes. The output display machine, equally error-free, converts this tape into negatives or positives, ready for offset or letterpress plate-making. Since this dream system makes no mistakes, there is no need for galley proofs or page proofs. This means no more printer's alteration charges, because authors will no longer get a chance to rewrite their manuscripts on galleys or page proofs.

Sure this sounds like a dream. But technologically, without too great an investment, we could make this dream come true today for practically any printing job, if we wanted to. Now I must sound a warning: Pushing technology too fast could change our dream into a nightmare. As many of you have noticed, quite a few parts of this dream are fuzzy. For instance, I didn't say whether these girls were in the publisher's office, a printing plant, a central service bureau, or a huge and drafty old castle in England. I didn't say where the computer was. Likewise, I didn't say who had the output machine, or what it was. All I did say is that we ended up with negatives or positives that were ready for conventional platemaking and printing.

Now let's see why we have to be so vague — why we need answers to so many questions before the ultimate role of computers in publishing can be pinpointed.

● System Design

The chief problem in computerized publishing is that there are so many different types of equipment and so many different techniques, each with advantages and drawbacks, for each part of our system. From these many variables, we must find the optimum system combination that will meet specific present and future publishing needs, reliably and economically. The glamour and publicity of new computer hardware must not lure us into premature action. This system design problem divides into four parts:

1. *Input hardware*, which converts the manuscript into machine-readable form.

[1] Manager, Information Research, McGraw-Hill, Inc., New York.

2. *Computer hardware,* which does the creative part of the printing process.
3. *Computer software,* including programs that provide for the production of proofs, correction of errors, and updating of subsequent editions.
4. *Output hardware,* consisting usually of a photocomposition machine that converts the computer output tapes to negatives or repro copy for printing.

• Input Hardware

A computer will accept editorial copy only in machine-readable form. This can be punched cards, punched tape, magnetic tape, or scanner-readable typed characters. Our manuscript must therefore be retyped, or rather keyboarded, first. Here we have three basic choices:

TAPE PERFORATORS

Most computer composition systems in operation today use tape-punching typewriters as input hardware. Flexowriters and Dura machines built around standard IBM electric typewriters are the most popular. Prices range from $1,500 to $2,500 each. There are also tape perforators without printing facilities, such as the Fairchild TTS perforators used in many printing plants. At least one of the machines should also have a tape reader to permit checking the performance of the machine by running punched tape through the reader and checking the hard copy produced from the tape.

The choice of tape-punching typewriters is difficult. On the Dura Mach 10 tape-punching typewriter, high typing speeds are possible because shift and unshift codes are punched automatically when the typist hits the conventional shift keys. On Friden Flexowriters each shift code must be punched separately by the operator. Unfortunately, some Dura machines get out of adjustment and occasionally lose a shift code when typists work at their maximum speed. Slowing down the typists eliminates this malfunction, but output is then down to that of Flexowriters. This leads to the conclusion that there is as yet no ideal tape-punching input machine for computer composition.

Punched paper tape for computers is often called idiot tape by printers, because it can be produced by ordinary typists without years of special training. This input tape contains nothing more than code equivalents of typed characters, without hyphenating or end-of-the-line indications, and with a minimum of special control characters for designating type font changes.

A typewriter keyboard has many advantages over the Linotype or Monotype keyboards currently used in printing plants. First of all, ordinary typists can be used for keyboarding of composition, after no more than a few hours of additional training. These typists can generally turn out at least 20% more work than is ordinarily obtained on keyboards of hot-metal casting machines. One reason is that a typist has only 44 keys to hit, while a Linotype operator has more than twice as many. Another reason is that typists do not have to slow up for the end-of-line justification or hyphenation decisions. The typewriter also gives hard copy immediately for proofreading.

KEYPUNCHES

Our second input possibility is a standard keypunch, driven by an essentially standard typewriter keyboard. The punched cards produced here can be imprinted either simultaneously with punching or in a separate machine to provide a line of printing across the top of the card for proofreading. Keypunching is viewed as an interim input measure for use with computers that have only punched-card input equipment, however, because it is slower and more expensive than punching paper tape.

SCANNERS

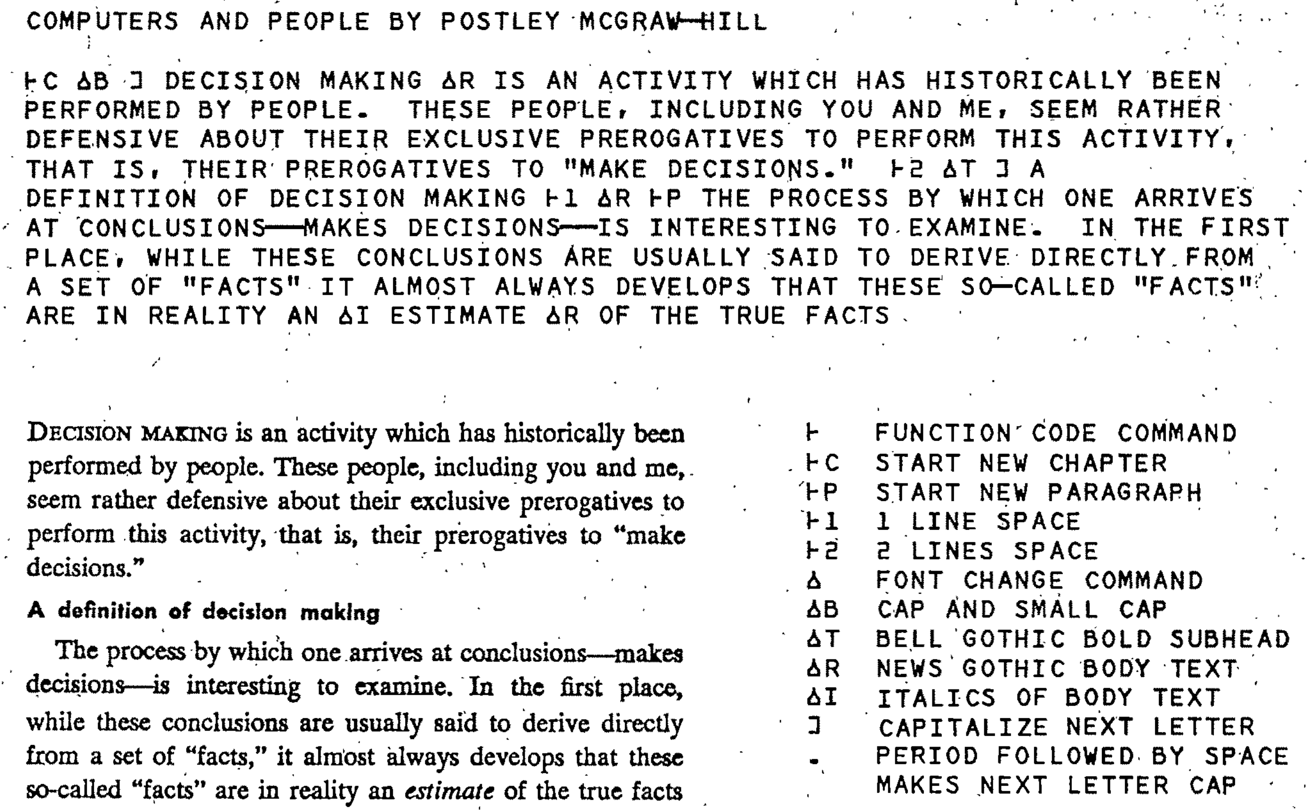
Controlled-font typewriters can be teamed up with a character-reading scanner for use as input hardware. Here editorial copy is retyped on an electric typewriter having a special font of all-caps characters than can be read accurately by a photoelectric scanning machine, such as the Farrington and Control Data Corp. scanners. Special controlled-font balls are now available for the IBM Selectric bouncing-ball typewriter.

Scanners generally deliver magnetic tape, ready for use as computer input. The scanner approach is attractive, but the all-caps limitation of the lower priced scanners is a drawback for input typing and for proofreading the hard copy.

An example of book composition as it might be typed for an all-caps scanner, in Fig. 1, shows how function codes might be used to obtain the desired typography. Note that the all-caps input lines do not end on the same words as the final typeset copy, because the computer ignores typewriter carriage returns. Note also that the computer has adjusted the spaces between words to achieve justification (lineup at the right). It could also hyphenate words whenever necessary to avoid excessive space between words.

It is possible today to purchase a scanner that will read both caps and lower case typing, in regular or controlled fonts. Cost is much more than for an all-caps scanner, however, and accuracy is still open to question. One of these, the Retina machine made by Recognition Equipment Corp., is being tested by Perry Publications in Florida for setting newspaper classified ads. Philco scanners have multi-font as well as lower case capability at correspondingly higher prices.

When a scanner for ordinary typing becomes available at a reasonable price, the pages of an author's manuscript could be fed directly into the scanner without retyping



```
COMPUTERS AND PEOPLE BY POSTLEY MCGRAW—HILL

⊦C ΔB ⊐ DECISION MAKING ΔR IS AN ACTIVITY WHICH HAS HISTORICALLY BEEN
PERFORMED BY PEOPLE• THESE PEOPLE, INCLUDING YOU AND ME, SEEM RATHER
DEFENSIVE ABOUT THEIR EXCLUSIVE PREROGATIVES TO PERFORM THIS ACTIVITY,
THAT IS, THEIR PREROGATIVES TO "MAKE DECISIONS•" ⊦2 ΔT ⊐ A
DEFINITION OF DECISION MAKING ⊦1 ΔR ⊦P THE PROCESS BY WHICH ONE ARRIVES
AT CONCLUSIONS——MAKES DECISIONS——IS INTERESTING TO EXAMINE• IN THE FIRST
PLACE, WHILE THESE CONCLUSIONS ARE USUALLY SAID TO DERIVE DIRECTLY FROM
A SET OF "FACTS" IT ALMOST ALWAYS DEVELOPS THAT THESE SO—CALLED "FACTS"
ARE IN REALITY AN ΔI ESTIMATE ΔR OF THE TRUE FACTS
```

DECISION MAKING is an activity which has historically been performed by people. These people, including you and me, seem rather defensive about their exclusive prerogatives to perform this activity, that is, their prerogatives to "make decisions."

A definition of decision making

The process by which one arrives at conclusions—makes decisions—is interesting to examine. In the first place, while these conclusions are usually said to derive directly from a set of "facts," it almost always develops that these so-called "facts" are in reality an *estimate* of the true facts

⊦	FUNCTION CODE COMMAND
⊦C	START NEW CHAPTER
⊦P	START NEW PARAGRAPH
⊦1	1 LINE SPACE
⊦2	2 LINES SPACE
Δ	FONT CHANGE COMMAND
ΔB	CAP AND SMALL CAP
ΔT	BELL GOTHIC BOLD SUBHEAD
ΔR	NEWS GOTHIC BODY TEXT
ΔI	ITALICS OF BODY TEXT
⊐	CAPITALIZE NEXT LETTER
•	PERIOD FOLLOWED BY SPACE MAKES NEXT LETTER CAP

FIG. 1. Example of book composition as typed for Farrington scanner (top), final output copy (lower left), and meanings of control code characters used by typist.

for conversion to magnetic tape. Colored pencil marks could be made at points where editing is desired. These would make the scanner read editing changes and corrections that have been typed between the lines or generate a code that tells the computer to take the desired change from another source.

● Computer Hardware

Many different general-purpose computers are suitable for electronic processing of manuscripts. The IBM 1620 general-purpose computer leads the picture here, with several dozen being used in printing plants as well as in newspaper applications. Software for newspaper composition on this computer is available from IBM. Next in popularity is the RCA 301, for which computer typesetting software has also been made available by the manufacturer. Other general-purpose computers used for composition include the Honeywell 200, Digital Equipment PDP-8, several Control Data models, the IBM 1400 series, and NCR 315. A few of these are used exclusively for composition, but in most installations the primary application is accounting. Composition work, for books in particular, is generally run during idle time on second or third shifts.

Choice of a particular general-purpose computer will generally be based on such factors as availability of the necessary operating time, availability of some or all the necessary software from the manufacturer, and availability of compatible input and output composition hardware for the computer. Size of the internal storage and speed of the input and output hardware are other factors to be considered. The output paper-tape punch speed is particularly important because it is generally operated on-line. A fast line printer is another asset; even better is one having a cap and lower case chain.

In general, the sum of the times required for the input reader, line printer, and output punch to handle their respective total characters will closely approach the total computer operating time for a particular composition job. The actual processing runs involving only magnetic tape handlers take a small percentage of the operating time on modern high-speed, general-purpose computers. This fact makes it possible to estimate composition run times with a high degree of accuracy, assuming availability of debugged programs, a good computer room operating staff, and an input that contains the correct control codes needed for proper processing by the computer.

Service bureaus offering computer composition include National Computer Analysts, Rocappi, and Documentation, Inc. These firms operate in competition with commercial printers because up to now only a few printers are using their computers for composition.

Special-purpose computers for composition include the Mergenthaler Linasec, Compugraphic DTP, three Harris-Intertype computers, and the Fairchild Comp/Set. These are intended chiefly for newspaper and printing plants, for use with tape-controlled hot-metal or photocomposition machines.

The most expensive of the three Harris-Intertype models provides a combination of logic and magnetic-drum dictionary lookup for automatic hyphenation. A simpler model uses only logic for hyphenation, so it occasionally inserts hyphens by guessing, when logic fails. The simplest and cheapest model stops automatically whenever it reaches a point where a word must be hyphenated to allow a human operator to make a decision on where the hyphen should go.

• Software Problems

A computer can only follow the instructions that are stored in its electronic memory. These instructions are known as software. Each program of instructions must allow for all possible combinations of input problems, yet must fit into the available core storage.

Computer software and computer hardware together serve to convert low-cost keyboarding by a typist to whatever is needed for driving the output composition machines. To do this, the software must include one or more of the following composition functions: (a) format and typography control; (b) justification; (c) hyphenation; (d) page makeup; (e) code conversion for the output hardware.

The control codes that specify line widths, indentations, type fonts, italicizing, bold-facing, subscripts, superscripts, leading (vertical spacing between lines), tabulating, and other parameters of composition must be converted by computer software into the units of length on which the computer bases its running width total of the characters already set in each line. To do this, the computer must be given beforehand the exact width of each character, in each of the fonts being used in the job being run. For justification, the computer must also be given the range of word spacing that will be permitted when justifying lines for a particular composition job. The greater the range, the less will be the need for hyphenating words. When the computer comes to a word that won't fit in the specified line width, it must reject that word and increase the spaces between words to fill the line.

A refinement of the justification program involves rearranging words in previous lines in a paragraph whenever a line cannot be justified within specified limits without hyphenating. The computer tries to transfer words from the end of one line to the start of the next, working back to the beginning of a paragraph, and staying within the word spacing limits, to see if a different arrangement of words can be achieved that will eliminate the need for hyphenating. This technique naturally works best for wide columns such as are used in books.

So far, software programs for computer composition have been written for specific jobs. This means that they must generally be revised rather extensively when changes in format, typography, or content are desired. At the present time, however, both RCA's Graphic Services Division and National Computer Analysts are working on programs that will be more or less universal, so as to permit reasonable changes in format. These programs are much more costly to produce, but will be cheaper in the long run because their costs can be spread out over a number of users of computer composition.

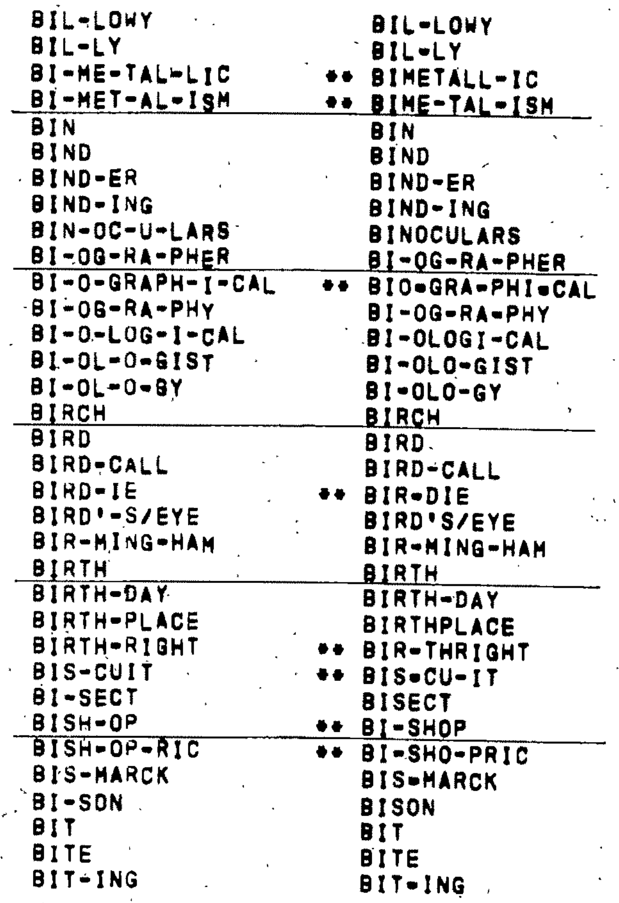
HYPHENATION

The software goal envisioned by many in computer-controlled typesetting is the introduction of hyphens correctly at ends of lines by means of a program that involves only logic, with no dictionary lookup of exception words. Perfection will never be achieved here as long as dictionaries are used as hyphenating guides, because proper names and many ordinary words in dictionaries do not follow logical rules for hyphenation. Scientific words derived from proper names (such as wattage, hyphenated watt-age because named after James Watt) are another headache, as also are words that are spelled the same but hyphenated differently depending on pronunciation. Worse yet, dictionaries do not agree on hyphenation. Even the Second and Third Editions of Merriam-Webster's Unabridged Dictionary differ from each other.

One program is available today that can equal or better the accuracy of human hyphenation while relying entirely on logic. This was written at National Computer Analysts in Princeton for an RCA 301 computer, and has been rewritten for Control Data and Univac computers. In developing this software, the computer was programmed to hyphenate test words in every possible position, compare its work with hyphenation as given in the Third Edition of Merriam-Webster Unabridged, and place asterisks ahead of each word having a hyphen in a wrong position. The resulting printout, part of which is shown in Fig. 2, was then used as a guide for further refining the hyphenating logic.

Here is a suggestion. Why not line up a stable of experts in linguistics, logic, typesetting, and proofreading to establish a consistent set of rules for hyphenation that are compatible with computer processing? A computer program could then be written from these rules to produce a word list showing the new and logical hyphenation for every word in the English language for publishing as the new standard hyphenating guide. Could we agree on a basic standard for hyphenation, based only on logic?

The majority of computer-controlled typesetting operations today are using a combination of logic and diction-



BIL-LOWY		BIL-LOWY
BIL-LY		BIL-LY
BI-ME-TAL-LIC	**	BIMETALL-IC
BI-MET-AL-ISM	**	BIME-TAL-ISM
BIN		BIN
BIND		BIND
BIND-ER		BIND-ER
BIND-ING		BIND-ING
BIN-OC-U-LARS		BINOCULARS
BI-OG-RA-PHER		BI-OG-RA-PHER
BI-O-GRAPH-I-CAL	**	BIO-GRA-PHI-CAL
BI-OG-RA-PHY		BI-OG-RA-PHY
BI-O-LOG-I-CAL		BI-OLOGI-CAL
BI-OL-O-GIST		BI-OLO-GIST
BI-OL-O-GY		BI-OLO-GY
BIRCH		BIRCH
BIRD		BIRD
BIRD-CALL		BIRD-CALL
BIRD-IE	**	BIR-DIE
BIRD'-S/EYE		BIRD'S/EYE
BIR-MING-HAM		BIR-MING-HAM
BIRTH		BIRTH
BIRTH-DAY		BIRTH-DAY
BIRTH-PLACE		BIRTHPLACE
BIRTH-RIGHT	**	BIR-THRIGHT
BIS-CUIT	**	BIS-CU-IT
BI-SECT		BISECT
BISH-OP	**	BI-SHOP
BISH-OP-RIC	**	BI-SHO-PRIC
BIS-MARCK		BIS-MARCK
BI-SON		BISON
BIT		BIT
BITE		BITE
BIT-ING		BIT-ING

FIG. 2. Early test of automatic hyphenation by logic alone. Left column shows hyphenation of Merriam-Webster Unabridged Third Edition and right column shows hyphenation by National Computer Analysts logic routines. Asterisks indicate misplaced hyphens. Hyphens omitted by computer were not counted as errors here, but newer test program counts both omitted and erroneous hyphens.

ary lookup for hyphenation. Here, when the computer reaches a word that must be hyphenated, it first looks in its internal dictionary of problem words that have been stored with hyphens in all possible positions. If the word is found there, the computer chooses a hyphenating position that places the line within the specified justifying limits. If the word is not in the internal dictionary, the computer then applies its stored rules of logic, based on common word endings and common combinations of letters, for placing the hyphen. If the word is one of the exceptions for which logic rules do not apply or have not yet been written into the program, the computer can arbitrarily place the hyphen after an odd-numbered letter, because statistics show that most hyphens fall after the 3rd, 5th, 7th, etc., letter in a word. In one program, when none of the logic rules applies, the computer in effect flips a coin and places the hyphen arbitrarily before or after a convenient vowel.

A third hyphenating alternative, which no one has tried as yet because it calls for more internal memory capacity than is available in any computer, would involve storing all of the words in the Merriam-Webster Unabridged Dictionary with all hyphen positions, for complete dictionary lookup. With a sufficiently large capacity in a random-access disk file or other large memory, it is not inconceivable that this approach may be tried in the future.

Finally, at the other extreme is manual hyphenation, wherein we have computer software that stops everything and sounds an alarm when a word is reached that requires hyphenation. An operator must then look at the word as shown on a computer display or as typed out by the computer and indicate where the hyphen should go. This manual operation is at present used only with small special-purpose computers, because it fails to utilize the hyphenating capabilities of a computer. On the other hand, with wide-column book composition an appropriate justifying program could reduce the need for hyphens to less than one line in 100. Here it may be more economical to use this combination of automatic and human procedures.

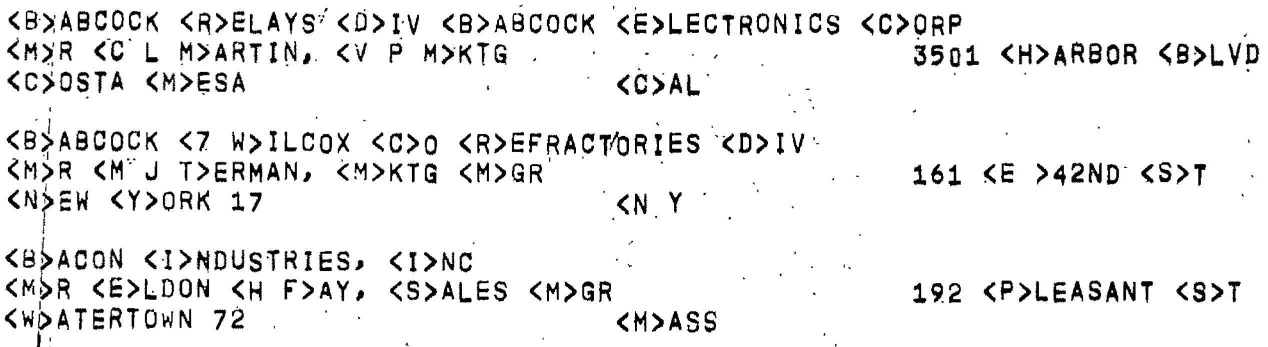
For page makeup of books, the available newspaper composition programs need to be expanded to cover handling of illustrations, computation of book length, and elimination of bad breaks such as widows, rivers, and heads at the bottom of a page or column. The preparation and testing of suitable programs for publishing requirements could be costly and time consuming for an individual publisher. Standard composition programs are now being prepared by some service bureaus for the benefit of all of their publisher clients.

ERROR DETECTION

IBM is already carrying out research on the use of a computer to detect errors in spelling and typing of editorial copy. One obstacle here is the size of the computer internal storage required for holding all of the words in the English language, with high-speed access to each word so the process does not appreciably slow up computer processing. Another drawback is that some typing errors create acceptable different words.

PHOTOCOMPOSITION CONTROL CODES

The last part of the software for computer-controlled typesetting is a program for converting the code format of the computer to that required by the composing machine and adding the necessary machine control codes. Computer manufacturers have developed this conversion software for the Teletypesetter tape required for hot-metal newspaper composition, and will presumably have conversion programs for photocomposition machines eventually.



```
<B>ABCOCK <R>ELAYS <D>IV <B>ABCOCK <E>LECTRONICS <C>ORP
<M>R <C L M>ARTIN, <V P M>KTG                             3501 <H>ARBOR <B>LVD
<C>OSTA <M>ESA                      <C>AL

<B>ABCOCK <7 W>ILCOX <C>O <R>EFRACTORIES <D>IV
<M>R <M J T>ERMAN, <M>KTG <M>GR                           161 <E >42ND <S>T
<N>EW <Y>ORK 17                     <N Y

<B>ACON <I>NDUSTRIES, <I>NC
<M>R <E>LDON <H F>AY, <S>ALES <M>GR                       192 <P>LEASANT <S>T
<W>ATERTOWN 72                      <M>ASS
```

Fig. 3. Use of lesser-than sign to indicate shift to caps and greater-than sign to indicate shift down to lower case as required for proofreading accurately from line printer proofs.

PROOFREADING

Normally, typeset material is proofread twice, on galley proofs and on page proofs. Computer composition will undoubtedly change this. Here are some of the possibilities that we must explore to determine the optimum proofreading procedures for each system from the standpoint of economics, accuracy, and acceptance by editors and authors.

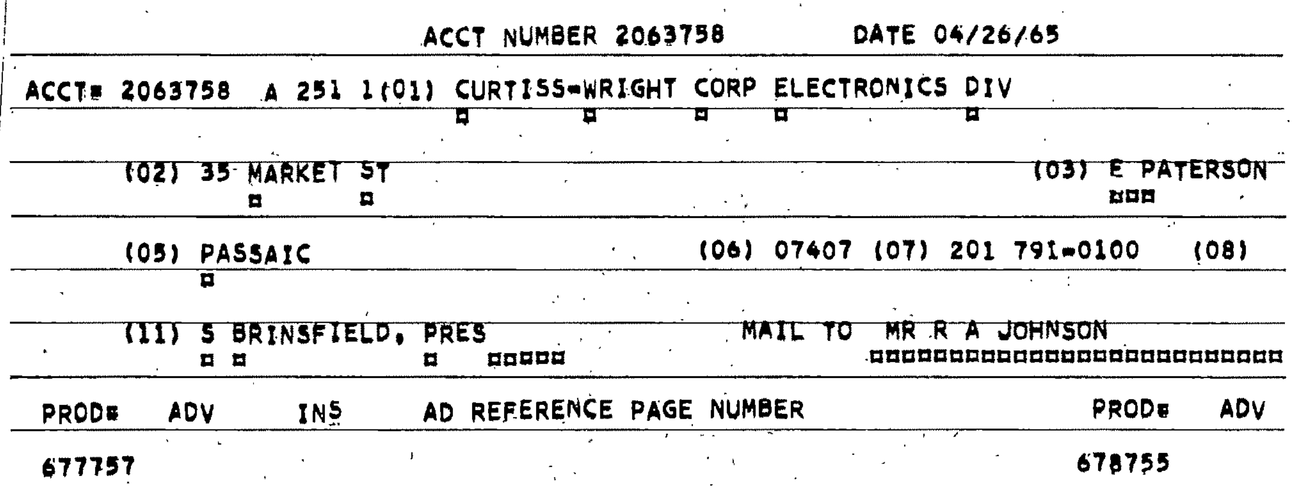
Proofreading of the hard copy produced on input tape-punching typewriters could be the only proofreading needed in an error-free computer system, if we could count on the typewriter and its punches to work perfectly. We can't yet. Furthermore, the hard copy at the input is not typed to final column width, hence there can be no checking of hyphenation, widows, and the other factors that determine high-quality composition. The hard input copy will be cap and lower case, however, and can have code characters to indicate italics, bold face, special characters, and other desired typographic changes.

The hard copy that is produced as a by-product of tape-punching is thus quite useful for detecting human errors in keyboarding. It will not show up machine errors occurring between the fingertips of the typist and the magnetic tape of the computer, however, so we need at least one more proofreading. How do we get it?

Use of a computer's own line printer to produce a display for proofreading is one logical answer. A major drawback, however, is that most line printers have only capital letters. Special programming is then needed to identify the letters that must be capitalized in the final output.

One method of indicating capitalization is based on the fact that three characters must be punched on paper tape each time a typist hits a capital letter. First comes the upper-shift character, produced when she touches the shift key. Next is the actual character, followed by the down-shift character that is punched when she releases the shift key. These shift characters can be converted to lesser-than and greater-than signs surrounding the letter or letters that are true caps on line printer proofs, as in Fig. 3. (Yes, we do get complaints from proofreaders with this approach.)

Another method involves more programming to remove the two shift characters from the printout and, instead, place an open lozenge or other special character on the line below, directly under each cap character, as in Fig. 4. This makes proofreading much easier but doubles line-



```
                         ACCT NUMBER 2063758        DATE 04/26/65

ACCT= 2063758  A 251 1(01) CURTISS=WRIGHT CORP ELECTRONICS DIV
                           □       □      □    □           □

      (02) 35 MARKET ST                                    (03) E PATERSON
              □      □                                          □□□

      (05) PASSAIC                     (06) 07407 (07) 201 791=0100   (08)
           □

      (11) S BRINSFIELD, PRES               MAIL TO  MR R A JOHNSON
           □ □           □    □□□□□        □□□□□□□□□□□□□□□□□□□□□□□□□□□

PROD=  ADV     INS     AD REFERENCE PAGE NUMBER                PROD=  ADV

677757                                                        678755
```

Fig. 4. Method of using open lozenges to indicate true capital letters on line printer proofs.

printer running time. This method is better from the standpoints of proofreading accuracy, proofreader morale, and labor vs. machine-time costs.

A cap-and-lower-case printer would give a more readable proof, but special characters would still be needed for italicizing, bold-facing, and characters not on the chain. The cost of this chain and the additional computer circuitry required must be weighed against the value of the lower-case printout for proofreading and its value in other applications. As one example, this cap-and-lower-case chain might be used to produce more readable repro copy for indexes at low cost compared to photocomposition or hot-metal typesetting, though at a serious penalty in characters per inch since the letters are uniformly spaced. Also, for a given level of readability, the cap-and-lower-case printout cannot be reduced in size as much as an all-caps printout and will therefore require more printed pages per job.

Proofreading of output photocomposition from Xerox copies of paper repro proofs or from blueprints of negatives is the ideal solution, though probably the most expensive. Here we see exact equivalents of conventional galley and page proofs showing all formats and all fonts of type called for.

CORRECTIONS

After errors have been caught by proofreading, paper tapes for the corrections must be punched with appropriate record, field, line, or other identifying codes. These correction tapes are fed into the computer for updating the master tape file, since this will be used for subsequent editions. Five procedures are now available, depending on the nature and amount of corrections:

1. Use the computer to punch output tapes only for lines requiring change, run these through the photocomposition machine, and get corrected lines for stripping into the negatives or pasting on the repro positives. This is generally the preferred procedure.
2. Use the computer to punch the entire output tapes over again for a second complete run through the photocomposition machines. This costly procedure will probably be preferable when there are drastic changes, such as might occur when a publisher allows an author to rewrite his manuscript on galley or page proofs.
3. Reperforating the punched paper input tape, with manual typing only when an erroneous word is reached, to obtain a perfect new tape for use as computer input. Chief drawback here is almost a doubling of input keyboarding time with no assurance that new errors won't be made.
4. Splicing of corrections into the original tape. This requires a person who can read punched characters, hence is rarely done.
5. Merging of correction tapes with the original tapes by operator switching of two readers. This takes even more input operator time than the third procedure and can give new errors if readers are not switched back and forth at the correct instants.

• Output Hardware

Output hardware includes everything needed to convert output magnetic tape of the computer to hot metal or to photographic negatives for plate-making. When the output machine requires punched paper tape, the computer itself or an off-line converter must drive a high-speed tape punch. With output machines like Photon ZIP and the forthcoming CBS-Mergenthaler Linotron, which will take magnetic tape directly, mag-to-paper conversion will not be needed. In commercial publishing, the punching of output paper tapes is generally done on-line, even though this means many more hours of computer time.

HOT-METAL OUTPUT

Many Linotype and Intertype casting machines will operate from punched paper tapes. The chief drawback to a hot-metal output machine, however, is the inherent mechanical error rate of the machine. In general, a machine malfunction will cause an error in about one out of every 50 cast slugs even with the best possible machine operation and maintenance. Until this error rate is eliminated, the trend in computer-controlled typesetting will logically be toward photocomposition machines, because these can theoretically operate without error.

PHOTOCOMPOSITION OUTPUT

Machines currently available for photocomposition from punched paper tape include those made by Photon, Harris-Intertype, ATF, Mergenthaler, and Alphatype. More are undoubtedly under development. These use a punched paper input tape to control the exposure of a film negative, character by character, with precise positioning of characters and precise control of spacing between lines, as determined by the codes produced during computer processing. The choice of a machine must be based on cost, speed, reliability, quality, the number of fonts of type required, the convenience of obtaining special characters, and the availability of backup machines if breakdowns occur.

ZIP

The high-speed Photon ZIP machine installed at the National Library of Medicine is the only machine that today takes computer magnetic tape directly. Its successful performance in producing Index Medicus is being closely studied by publishers and printers. The price tag today is $200,000, with the user furnishing his own computer. This first ZIP is setting about 300 characters per second, which means that it can expose all the negatives for a 100,000-word book in less than an hour. This is equal to about 30 Linotypes. ZIP holds 264 characters at a time, and these can be changed by sliding in a new set of glass matrices.

The quality of the output of ZIP is satisfactory for indexes and directories, but most book and magazine publishers require better work. Chief defects are erratic vertical positioning of letters and variations in letter density.

Each line on the negatives for Index Medicus is exposed across all three columns, because their computer also does page makeup. The internal memory or external disc file must therefore have enough capacity to hold the contents of an entire page, in addition to the required working storage capacity.

The second model of ZIP is being used experimentally by Western Electric Co. for printing daily and monthly changes in some New York City telephone directories, for use by operators. The third is scheduled to go to England. There is real hope that an improved ZIP will take the magnetic output tapes of the computer directly and give graphic arts quality in a variety of type fonts.

With a high-speed photocomposition machine that accepts magnetic tape directly, there may be less of a cost penalty for using the machine to produce proofs. The machine is so fast that it would likely be idle part of the time on second or third shifts, if not on the first shift. It is then logical to use the machine for producing either paper prints or film negatives for proofreading, since the out-of-pocket extra cost is only for the photographic proof paper.

PHOTON 713

A few of the new Photon 713 photocomposition machines are now in the field. Early reports indicate that this machine has real promise for producing composition of graphic arts quality from punched paper tapes of computers. The specifications definitely offer advantages over other photocomposition machines available today. Speed is about 20 characters per second, with a choice of 720 characters or eight full fonts in eight type sizes, in producing either negatives or repro positives from punched paper tape. Price is about $50,000. Photon is also offering, for $15,000 additional, the same capability working directly from the magnetic output tapes of a computer. This version, if it performs reliably, will be of high interest to printers and service bureaus because it eliminates the slow and costly extra step of using a computer to convert magnetic tape to punched paper output tape.

FOTOTRONIC

Another promising photocomposition machine for punched paper output tapes of computers is the Harris-Intertype Fototronic, now selling for around $55,000. This has the same rated speed as the Photon 713, but actual throughput varies with the number of type size and font changes required on a given job.

CATHODE-RAY PHOTOCOMPOSITION

Many cathode-ray character-generating systems have been proposed for direct operation from magnetic tape, but as yet none is on the market in commercial form. The Government Printing Office awarded jointly to Mergenthaler Linotype Company and CBS Laboratories a $2,185,000 contract to produce two such machines, called Linotrons, for 1966 delivery. Speed varies with type size, because it takes longer to make an electron beam create a larger character from a pattern of fine lines, but is expected to be much faster than ZIP. For making paper proofs, which can be lower in quality as long as they are readable, the Linotron can run at up to 5,000 characters per second. Linotron will give a choice of 256 characters in one basic font that can be changed electronically to any of eight different sizes ranging from 5 to 18 points. Resolution is claimed to be better than Linotype hot-metal work but not quite up to Linofilm photocomposition quality. A price of $500,000 is being quoted for a commercial model.

Other firms working on the cathode-ray approach include the RCA Graphic Services Division, Alphanumerics, K. S. Paul & Associates in England, and Rudolf Hell in Germany.

The role of costly, high-speed cathode-ray machines in commercial publishing is as yet unknown. Smaller and slower machines can share the work and back up each other. A fast machine is economically unsound unless it can be kept busy or can be justified on the basis of its high speed for a few specific jobs.

LINE PRINTER OUTPUT

With proper adjustment and operation of a high-speed line printer, using a high-quality new nylon ribbon and a high grade of paper, repro copy can be obtained directly from a computer. This can then be reduced in size photographically to produce plates for offset printing of indexes, directories, and other types of reference books. Alternatively, paper masters can be produced directly on the line printer if the original large type size is acceptable.

The two chief drawbacks of line printers are the waste of space inherent in uniform spacing of characters on a line printer and the objections of users to the all-caps print-out of most line printers. The cap-and-lower-case chain that is available for the IBM 1403 line printer gives improved readability in the new typewriter-like font, but the spacing is still the same. In one case, for Index Medicus, this cap-and-lower-case printer is backup for the Photon ZIP, but an emergency issue produced on the line-printer will take twice as many pages.

The line printer is envisioned only for jobs where less than graphic arts quality can be tolerated and space requirements are not critical. Indexes, book-form library catalogs, parts lists, and some directories are examples of work for which a line printer should be considered.

● Economical Applications

Although real progress has been made in computer composition, there are today very few economical commercial applications that provide graphic arts quality. One reason for this has been the lack of photocomposition machines suitable for computer composition. Another is the lack of completely versatile software.

Production costs drop most spectacularly when printing jobs make maximum use of the data processing capabilities of computers, as with indexes and directories. Here are some job characteristics that favor use of computers:

1. A need to explode information so a given item of input is duplicated many times by the computer.
2. A need to sort items of information, so they appear in one or more desired sequences regardless of the order in which the items enter the computer.
3. A need for updating the information, by cumulating corrections and new material with old material, so the computer can be used to eliminate rekeyboarding of the unchanged older material.
4. A need for speed that overrides cost considerations.

One publication that met these characteristics was the annual Electronics Buyers' Guide. This McGraw-Hill publication tells who makes the components and equipment that constitute the electronic industry. The input data for this directory is the equivalent of 300 pages in print, while the output or final directory is almost 800 pages in print — a tremendous explosion of information. There is a real requirement for data processing here also; the input from questionnaires requires three alpha sorts. Just before the cutoff deadline for input, the entries of advertisers must be changed to bold face type and sequenced separately. And finally, there is the requirement for updating once a year by incorporating any changes in the manufacturer's name and address, phone number, corporate data, addresses of representatives, or products made. The potential for saving by computer was so attractive that a decision was made to produce it this way for the first time in 1965.

Since there was no precedent for producing a directory of this size by computer, many decisions had to be made in conjunction with an exhaustive systems study. While describing the procedures finally adopted, the problems and options will be covered for guidance in planning similar jobs. Here are the main steps:

1. *Assign Code Numbers.* First of all, the 7,000 manufacturer names were arranged manually in the desired alphabetic sequence and fed into a computer for automatic assignment of 6-digit manufacturer code numbers, with a spacing of 125 between the numbers. This permits insertion of new names in correct alphabetic sequence in future years. Each new name is given a number halfway between the numbers of adjacent names to leave room for inserting further names. Product headings were similarly tape-punched and given 5-digit codes.

When assigning codes, the computer also adds a check digit at the end of each number to permit automatic computer detection of errors in typing of code numbers. If an error is made in a code number, the computer will get a different check digit and thus detect the errors.

2. *Prepare Typographic Specifications.* This is a much bigger job than it sounds. Rules must be established for each detail of type size, style, special characters, punctuation, indents, capitalization, handling of turnover lines, column widths, spacings between lines, leader dots, rules for breaking an address when it will not all go on one line, etc. The entire input must be divided into logical fields, and the maximum number of characters in each field must be specified. Fields were made as small as possible; examples of fields are the corporate name, street address, city, state, ZIP code, phone number, and number of engineers employed. The importance of establishing specs and field lengths beforehand cannot be emphasized too much, because changes in computer programming can be very costly.

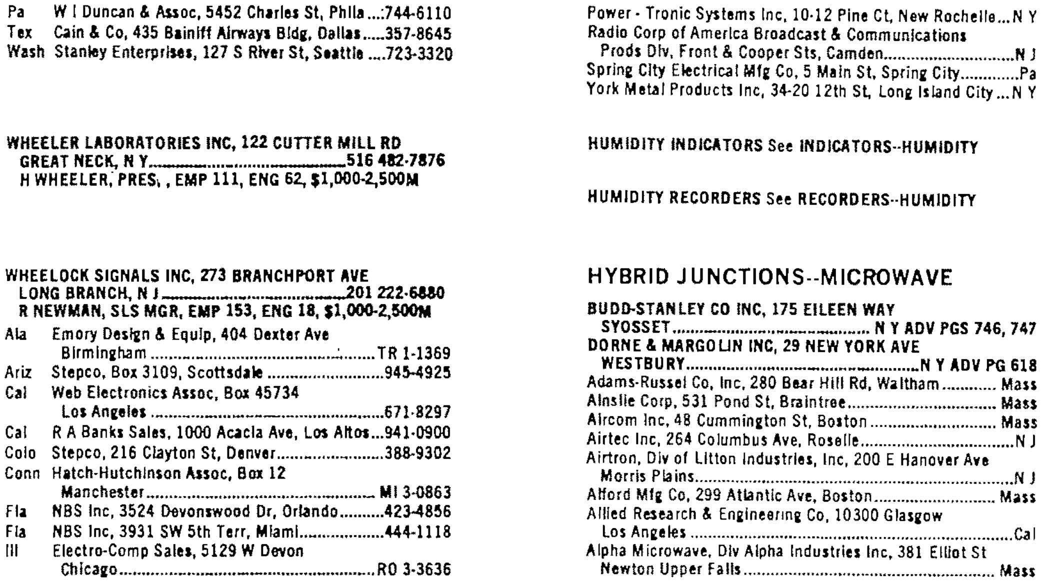
The column width was 200 units of 6-point type, which came out to be 14.1527 picas. Character widths were specified to thirds of a unit, so each line was made up of 600 increments. The width of each character in increments had to be determined and fed into the computer first.

Instead of justifying right-hand margins of columns by changing word spaces and hyphenating, the first lines of an entry were left short when the next field wouldn't fit. For the last line of each entry, leader dots were specified to make the phone number or state abbreviation come out flush right. This illusion of justification does not look too bad, as can be seen in Fig. 5.

3. *Choose a Computer.* The decision here went to a 24K Honeywell 200 computer in the McGraw-Hill Data Processing Center in Hightstown, New Jersey, used with a Honeywell 500-character-per-second paper tape reader and a 100-character-per-second paper tape punch.

4. *Choose Output Machines.* Here the choice was the American Type Founders B-8 paper-tape-driven photocomposition machine, which is reliable even though painfully slow (about as fast as an average typist, at 5 characters per second). The slowness is an asset, however, because the four machines required to meet the production schedule provided backup for each other in the event of trouble. Special 176-character type discs had to be designed and made to get the variety of type fonts and sizes required (6-point c & lc, 6-point b-f caps, and 8-point b-f caps).

5. *Select Programmers.* Programming was contracted to National Computer Analysts in Princeton, New Jersey.



```
Pa   W I Duncan & Assoc, 5452 Charles St, Phila ...:744-6110
Tex  Cain & Co, 435 Bainiff Airways Bldg, Dallas.....357-8645
Wash Stanley Enterprises, 127 S River St, Seattle ....723-3320


WHEELER LABORATORIES INC, 122 CUTTER MILL RD
  GREAT NECK, N Y.................................516 482-7876
  H WHEELER, PRES, , EMP 111, ENG 62, $1,000-2,500M


WHEELOCK SIGNALS INC, 273 BRANCHPORT AVE
  LONG BRANCH, N J ...............................201 222-6880
  R NEWMAN, SLS MGR, EMP 153, ENG 18, $1,000-2,500M
Ala  Emory Design & Equip, 404 Dexter Ave
       Birmingham .....................................TR 1-1369
Ariz Stepco, Box 3109, Scottsdale ........................945-4925
Cal  Web Electronics Assoc, Box 45734
       Los Angeles ......................................671-8297
Cal  R A Banks Sales, 1000 Acacia Ave, Los Altos...941-0900
Colo Stepco, 216 Clayton St, Denver.......................388-9302
Conn Hatch-Hutchinson Assoc, Box 12
       Manchester....................................MI 3-0863
Fla  NBS Inc, 3524 Devonswood Dr, Orlando..........423-4856
Fla  NBS Inc, 3931 SW 5th Terr, Miami..................444-1118
Ill  Electro-Comp Sales, 5129 W Devon
       Chicago.........................................RO 3-3636
```

```
Power - Tronic Systems Inc, 10-12 Pine Ct, New Rochelle...N Y
Radio Corp of America Broadcast & Communications
  Prods Div, Front & Cooper Sts, Camden..........................N J
Spring City Electrical Mfg Co, 5 Main St, Spring City............Pa
York Metal Products Inc, 34-20 12th St, Long Island City...N Y


HUMIDITY INDICATORS See INDICATORS--HUMIDITY


HUMIDITY RECORDERS See RECORDERS--HUMIDITY


HYBRID JUNCTIONS--MICROWAVE
BUDD-STANLEY CO INC, 175 EILEEN WAY
  SYOSSET.......................................N Y ADV PGS 746, 747
DORNE & MARGOLIN INC, 29 NEW YORK AVE
  WESTBURY.............................................N Y ADV PG 618
Adams-Russel Co, Inc, 280 Bear Hill Rd, Waltham ........... Mass
Ainslie Corp, 531 Pond St, Braintree............................ Mass
Aircom Inc, 48 Cummington St, Boston .......................... Mass
Airtec Inc, 264 Columbus Ave, Roselle...............................N J
Airtron, Div of Litton Industries, Inc, 200 E Hanover Ave
  Morris Plains.........................................................N J
Alford Mfg Co, 299 Atlantic Ave, Boston......................... Mass
Allied Research & Engineering Co, 10300 Glasgow
  Los Angeles ..........................................................Cal
Alpha Microwave, Div Alpha Industries Inc, 381 Elliot St
  Newton Upper Falls................................................ Mass
```

FIG. 5. Typography of Manufacturers Section (left) and Product Section (right) of 1965 Electronics Buyers' Guide, as produced by running computer-produced paper tapes through ATF model B-8 photocomposition machines. Computer is programmed to insert exactly the correct number of leader dots ahead of phone number and state fields to make these line up at right, giving effect of justification.

This firm was chosen chiefly because their programmers had heavy experience in computer composition.

6. *Punch Input Paper Tapes.* This input punching was farmed out to the programming firm to achieve single responsibility for establishing and punching the necessary control codes. Here it was found that the hard copy did not always correspond to the punches made by the two Dura Mach 10 tape-punching typewriters. Accordingly, the tapes were fed back into the readers of the Duras to produce new hard copy for proofreading against the questionnaires. Error correction required very little additional keyboarding, because only the manufacturer code, the field number, and the new corrected wording for the field had to be punched. The average length of a directory field is only about three words.

7. *Produce Line Printer Proofs.* Batches of the punched tape were fed into the computer along with the correction tapes for converting to magnetic tape, merging corrections, sorting records into final sequence, making validity checks of code numbers, counting characters to make sure no field was longer than the maximum length provided for it, and printing proofs. Since the line printer could print only capital letters, some means of identifying true capital letters had to be used. The decision to print an open lozenge under each true capital letter, as in Fig. 4, worked out very well from the proofreading standpoint, even though it doubled the line printer running time.

8. *Punch Output Paper Tape.* After coded instructions for bold-facing of advertisers had been punched and inputted, the computer selected the fields of data needed, processed these as required for the final sequences in the two sections of the directory, and added the necessary control codes for the photocomposition machines. Output tapes were then punched for the Manufacturer's Section in 21 hours of Honeywell 200 computer time and for the Product Section in 35 hours, to give a total of about 28 miles of 8-channel paper tape.

9. *Convert Tapes to Repro Positives.* The output of paper tape was run through three ATF B-8 machines to get paper prints in about 20 days of three shift operation. (An additional machine was kept in the publisher's plant for making last minute corrections and to serve as backup for those in the printer's plant.) The paper repro prints were dummied along with ad proofs, then photographed to get page negatives for making offset printing plates.

Next Edition. The magnetic tape reels containing the input data were stored, for two purposes: (a) to produce individual questionnaires for acquiring data for the next edition, with addresses in correct positions for window envelopes; (b) to repeat unchanged material in the next edition without additional keyboarding. Customized questionnaires make it easy for manufacturers to check what they had in the previous directory and mark the changes desired. It is estimated that only about 50 pages of new input will be keyboarded in 1966 to get 800 pages of output.

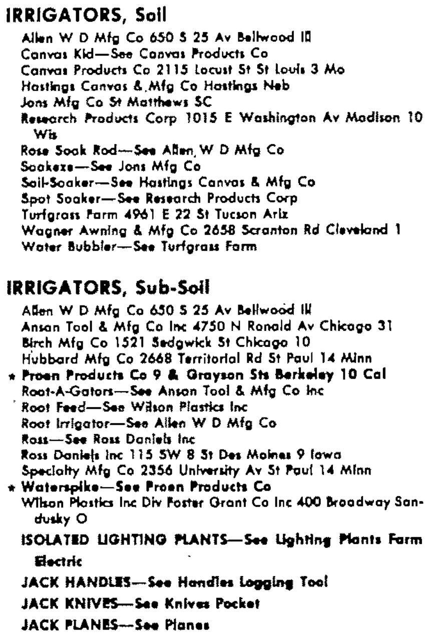
Other Directories. Chilton's Hardware Age directory was produced on a computer in 1965, using the facilities of Rocappi in Philadelphia. The hardware consisted of Dura Mach 10 tape-punching input typewriters, an RCA 301 computer, and a Photon 513 output photocomposition machine. The end product was 329 pages of composition in 6-point type, having the typographic format of Fig. 6.

• Book Catalogs for Libraries

Computer composition techniques have made it possible for many libraries to replace their catalog card files with much more convenient book-type catalogs. These are usually updated annually, and cumulative supplements for new material are issued quarterly. The catalogs can serve a number of main libraries as well as all branch locations.

One example is the combination catalog serving the medical libraries of Yale, Harvard, and Columbia, produced by computer under the direction of F. G. Kilgour of Yale.

Another significant example is being produced by Documentation, Inc., for the Baltimore County Public Library. There are three hardbound annual volumes, for Title, Author, and Subject, containing 1,500, 1,500, and 1,700 pages respectively to cover 50,000 titles. Press run is about 100 sets. Here the cap and lower case printout of an IBM 1403 chain printer is reduced photographically to 8-point for offset printing to give the format shown in Fig. 7. Punched cards are used for input. Processing is done on an IBM 1401, as also are sorting, breakout, and cumulating of the paper cover quarterly supplements.



IRRIGATORS, Soil

Allen W D Mfg Co 650 S 25 Av Bellwood Ill
Canvas Kid—See Canvas Products Co
Canvas Products Co 2115 Locust St St Louis 3 Mo
Hastings Canvas & Mfg Co Hastings Neb
Jons Mfg Co St Matthews SC
Research Products Corp 1015 E Washington Av Madison 10 Wis
Rose Soak Rod—See Allen W D Mfg Co
Soakeze—See Jons Mfg Co
Soil-Soaker—See Hastings Canvas & Mfg Co
Spot Soaker—See Research Products Corp
Turfgrass Farm 4961 E 22 St Tucson Ariz
Wagner Awning & Mfg Co 2658 Scranton Rd Cleveland 1
Water Bubbler—See Turfgrass Farm

IRRIGATORS, Sub-Soil

Allen W D Mfg Co 650 S 25 Av Bellwood Ill
Anson Tool & Mfg Co Inc 4750 N Ronald Av Chicago 31
Birch Mfg Co 1521 Sedgwick St Chicago 10
Hubbard Mfg Co 2668 Territorial Rd St Paul 14 Minn
* **Proen Products Co 9 & Grayson Sts Berkeley 10 Cal**
Root-A-Gators—See Anson Tool & Mfg Co Inc
Root Feed—See Wilson Plastics Inc
Root Irrigator—See Allen W D Mfg Co
Ross—See Ross Daniels Inc
Ross Daniels Inc 115 SW 8 St Des Moines 9 Iowa
Specialty Mfg Co 2356 University Av St Paul 14 Minn
* **Waterspike—See Proen Products Co**
Wilson Plastics Inc Div Foster Grant Co Inc 400 Broadway Sandusky O

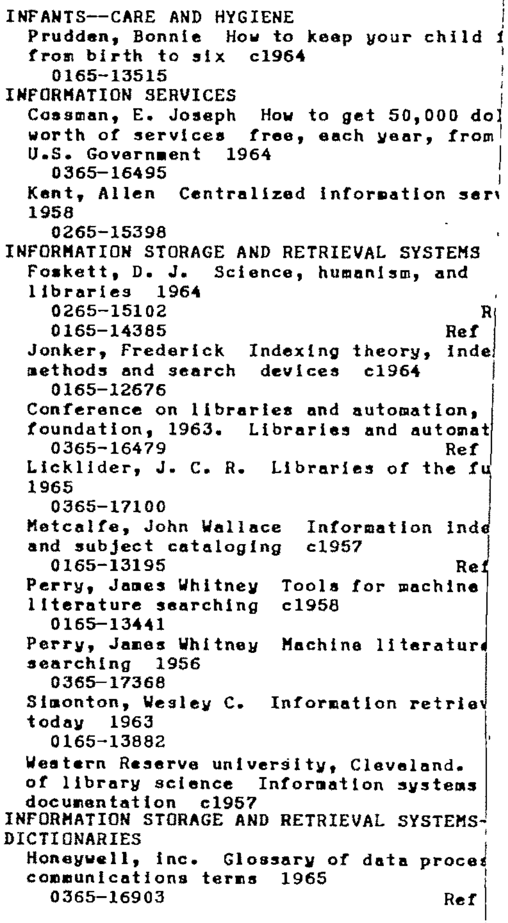
ISOLATED LIGHTING PLANTS—See Lighting Plants Farm Electric

JACK HANDLES—See Handles Logging Tool

JACK KNIVES—See Knives Pocket

JACK PLANES—See Planes

Fig. 6. Typography of Hardware Age as produced by a Photon 513 from computer-processed information. Bold-faced lines preceded by star indicate advertisers. Cross-references for trade names are alphabetized in same sequence with manufacturers. This directory has only a product section.



```
INFANTS--CARE AND HYGIENE
  Prudden, Bonnie  How to keep your child f
  from birth to six  c1964
    0165-13515
INFORMATION SERVICES
  Cossman, E. Joseph  How to get 50,000 dol
  worth of services  free, each year, from
  U.S. Government  1964
    0365-16495
  Kent, Allen  Centralized information ser
  1958
    0265-15398
INFORMATION STORAGE AND RETRIEVAL SYSTEMS
  Foskett, D. J.  Science, humanism, and
  libraries  1964
    0265-15102                          R
    0165-14385                        Ref
  Jonker, Frederick  Indexing theory, inde
  methods and search  devices  c1964
    0165-12676
  Conference on libraries and automation,
  foundation, 1963.  Libraries and automat
    0365-16479                        Ref
  Licklider, J. C. R.  Libraries of the fu
  1965
    0365-17100
  Metcalfe, John Wallace  Information inde
  and subject cataloging  c1957
    0165-13195                         Re
  Perry, James Whitney  Tools for machine
  literature searching  c1958
    0165-13441
  Perry, James Whitney  Machine literatur
  searching  1956
    0365-17368
  Simonton, Wesley C.  Information retrie
  today  1963
    0165-13882
  Western Reserve university, Cleveland.
  of library science  Information systems
  documentation  c1957
INFORMATION STORAGE AND RETRIEVAL SYSTEMS-
DICTIONARIES
  Honeywell, inc.  Glossary of data proce
  communications terms  1965
    0365-16903                        Ref
```

Fig. 7. Portion of October 1965 supplement book catalog that now replaces catalog card fi more County Public Library system, as produce 1403 line printer having cap and lower case chain

Other computer-processed book catalogs in of the University of Toronto Library and Atlantic University Library. A number of dustrial libraries are using the same computer but using the line printer printouts directly because the three or four copies produced in t adequate for their needs.

• Summary

Computer composition is economical today, types of directories, indexes, cumulated bib and other works involving significant amount and other manipulation of input information.

For straight text that requires only justific .nd hyphenation, it is much more difficult to promi cost savings at this time. Here also there is a major technological problem — the inability to produce at computer speeds a printout that will be acceptable to proofreaders, editors, and authors in place of conventional printed galley proofs. The cathode-ray approach to character generation does offer promise here as a proof printer. Cost and speed of the machine should approximate that of a line printer, but it must be able to give easily readable proof prints in a wide variety of type faces, sizes, and special characters.

Another important deterrent to widespread adoption of computer composition is the high cost of software. Work on basic compiler programs having sufficient flexibility to handle different jobs without reprogramming is now under way. Costs are high, well up into six figures. If these programs can be made sufficiently universal to permit amortizing cost over a large number of jobs, the number of economical applications for computer composition should increase tremendously.

Input hardware problems are being solved at a satisfactory rate with a variety of keyboard units that punch paper tape with or without hard copy. Some keyboard units even produce magnetic tape, but the higher cost per machine may preclude widespread use.

The output photocomposition hardware picture also looks more encouraging in 1966, now that Photon 713's and Harris-Intertype Fototronics are in actual use in printing plants. These machines sell for about four times the price of an ATF B-8, but provide a greater variety of type fonts and sizes along with higher operating speeds.

Cathode-ray photocomposition machines are appearing also this year in Europe as well as this country, but it remains to be seen whether they can consistently and reliably provide the graphic arts quality required by most book and magazine publishers.

Bibliography

1. Austin, Charles J. 1965. The MEDLARS Project at the National Library of Medicine. *Libr. Resources & Tech. Ser.,* **9:** 94–99. Winter issue.
2. Quinn, Hugh J. 1965. Computer Magic. *Printing Magazine/National Lithographer,* **89** (11): 42–44, 47, 64. Directory production.
3. *Roadblocks to Computer Composition; Editing and Proofreading for Computer-Processed Books.* 1965. *Book Production Industry,* **41** (9): 58–63.
4. Hard Copy Proofs; What Authors Need to Know to Work with Hard Copy. 1965. *Book Production Industry,* **41** (6): 53–57.
5. Computers in Composition 1965; The Debate Grows on Hyphenation. 1965. *Book Production Industry,* **41** (4): 53–59.
6. Santarelli, P. F. *Computer Prepared Text: A Real-Time/Time-Sharing Multi-Terminal Publication System.* IBM Technical Report TR 00.1263, April 20, 1965. Poughkeepsie: Systems Development Division. 36 pages.
7. Hattery, Lowell H., and Bush, George P. 1965. *Automation and Electronics in Publishing.* Washington: Spartan. 206 pages. Sixteen papers from American University 1965 symposium, plus 211-item bibliography.
8. Strauss, Victor. 1965. *The Printing Industry.* Washington: Printing Industries of America. Chapter II, Section 6 contains 32 pages dealing specifically with computerized composition.
9. *Proceedings of International Conference on Computerized Typesetting,* March 2–3, 1965. Washington: Research and Engineering Council of the Graphic Arts Industry, Inc. 157 pages. Nineteen papers plus discussions.
10. Computerized Typesetting: Interest Runs High. *Publisher's Weekly,* April 5, 1965, pp. 52–68. Report on Mar. 1965 Research and Engineering Council Conference on Computerized Typesetting.
11. Mathews, M. V., and Miller, Joan E. 1965. Computer Editing and Image Generation. *AFIPS Conference Proceedings — Fall Joint Computer Conference,* **1:** 389–398. Washington: Spartan.
12. *Proceedings of Computer Typesetting Conference,* London University, 1964. 1965. London: Institute of Printing Limited. 245 pages.
13. New Equipment and Trends in Automated Composition. 1964. *Book Production Magazine.* **80** (12): 36–39.
14. Bennett, David. The Case for Unjustified Typsetting. *British Printer,* October 1964. 5 pages. Reprinted by Composition Information Services, 1605 N. Cahuenga Blvd., Los Angeles.
15. Gardner, Arthur E. The Age of Computerized Typesetting . . . Phase 2. 1964. *Printing Production,* **95** (10): 48–53.
16. Getting Started in Computer Composition. 1964. *Book Production Magazine,* **80** (9): 52–54.
17. Weinstein, Edward A., and Spry, Joan. 1964. Boeing SLIP: Computer Produced and Maintained Printed Book Catalogs. *Am. Doc.,* **15:** 185–190.
18. Ohringer, Lee. *Computer Input from Printing Control Tapes.* A paper presented at the 16th Annual Meeting of the Technical Association of the Graphic Arts, Pittsburgh, Pa. June 3, 1964. 13 pages.
19. Barnett, Michael P., Moss, D. J., and Luce, D. A. 1964. Computer Generation of Photocopying Control Tapes. II. The P C 6 System. *Am. Doc.,* **15:** 115–120.
20. The (R)evolution in Book Composition. Part 3: What's Ahead for Computers; Part 4: The Systems Concept — Key to Computer Profits, Management and the Computerized Future. *Book Production Magazine,* **79:** 55–61, (April 1964) and 67–73 (May 1964.
21. Seybold, John W. 1964. The ROCAPPI System for Computerized Composition. *Book Industry Magazine,* **1** (3): 42–45.
22. Holliday, Alan S. Computer Controlled Composition for Books. Part 1: Concepts, Systems, Machines, Manning, Problems, and Solutions; Part 2: Applying the RCA 301 Computer to Book Typesetting. *Book Industry,* **1:** 22–25 (Feb. 1964) and 28–31, 78. (Mar. 1964).
23. *Computers: Their Impact on Book Composition.* Special 32-page report reprinted from Feb. 1964 *Book Production Magazine,* containing five articles: "Computers in '64: Year of Transition from Theory to Practice"; "Kingsport and Computers: A Book Manufacturer's Experience in Composition Research"; "What's Ahead

for Computers?"; "Computers Are Here-What Now?"; "The Systems Concept — Key to Computer Profits."

24. An Introduction to Computer Typesetting. Part 1: Basic Computer Principles; Part 2: The Automation of Typesetting in Application. *Print in Britain,* **11:** 20–22. (Jan. 1964) and **11:** 27–32 (Feb. 1964).
25. Buckland, Lawrence F. The Recording of Library of Congress Bibliographic Data in Machine Form. Maynard, Mass.: Inforonics, Inc. 43 pages.
26. Duncan, C. J. 1964. Look! No Hands. *Penrose Annual,* **57:** 121–167.
27. Typesetting in the Computer Age. 1964. *Print in Britain,* **12:** 8-page supplement.
28. Gardner, Arthur E. 1964. Computerized Typesetting — A Management Report on the State of the Art. 11 pages. *Composition Information Services Newsletter,* Los Angeles, Calif.
29. Duncan, C. J., Molyneux, Eve L., Page, E. S., and Robson, M. G. 1963. Computer Typesetting: An Evaluation of the Problems. *Printing Technology,* 133–151 (Dec. 1963).
30. Bozman, William R. 1963. Phototypesetting of Computer Input. NBS Technical Note 170. Washington: U. S. National Bureau of Standards. 6 pages.
31. Barnett, Michael P., and Kelley, K. L. Computer Editing of Verbal Texts. Part 1. The ESI System. *Am. Doc.,* **14:** 99–108 (April 1963); **15** (2): 115–120 (April 1964).
32. Smith, Frank H. 1963. Computers and Composition. *Mod. Lithographer,* **31** (1): 37–44.
33. North, Arthur. 1963. Quality Typography from Computer Data. 12-page booklet. Washington: U. S. Patent Office. Covers computer conversion of all-caps input of directory names and addresses to cap and lower case output.

Cost Distribution and Analysis in Computer Storage and Retrieval[1]

A method for costing computer jobs done by a mechanized storage and retrieval activity is proposed and discussed. Attention is confined solely to computer costs. The Science Information Exchange, a mechanized installation handling information on research in progress is used as the case in point. All computer jobs are grouped as batched, singly run or maintenance tasks. Job unit costs are calculated with and without inclusion of file maintenance costs. Should other activities compute their costs similarly, inter-activity cost comparisons can be made readily, opening the door to cost-quality criteria for mechanized searches and report preparation.

HARVEY MARRON[2] and MARTIN SNYDERMAN, Jr.[3]

• Introduction

Despite the widespread interest in the economics of computer storage and retrieval of scientific information — specifically the cost of performing searches and preparing reports — there is little definitive discussion of this subject in open literature. A few operating installations have released data on unit search costs, usually in terms of computer time per job obtained by dividing total computer processing time for a batch of jobs by the number of jobs. By doing so, however, the costs for separate jobs or building, modifying, maintaining, and updating the file are neglected. Depending on file array, maintenance procedures, and the file update frequency, these costs may be large and can rarely be disregarded. It can be shown that this is especially true if the search files are updated often but seldom searched.

This paper is confined solely to a technique for distributing computer costs. This is not to say that other associated costs, such as initial file design, programming, or general administrative overhead are unimportant. Quite the contrary. They are very important and in the last analysis must be brought into the computation for the true reflection of costs. However, the distribution of costs is complex and relatively uncharted and in our opinion the building block approach to a final solution in this area is preferable to an over-all assault on the total problem.

There are two aspects of computer jobs costs which must be considered: (1) the direct costs of doing a specific job and (2) the indirect costs of maintaining (not to be confused with the initial building) the files which are used in doing these jobs. Both aspects will be discussed in this paper.

• System Description

The Science Information Exchange (SIE) is a clearinghouse for current research in the life, physical, and social sciences. Information within its scope of coverage is collected, indexed, stored, and made available on demand to interested members of the scientific community. Requestors for information have ranged from bench level scientists to highly placed scientific program managers/administrators.

Essentially, information on six aspects of each research project is acquired, indexed, and stored. These are:

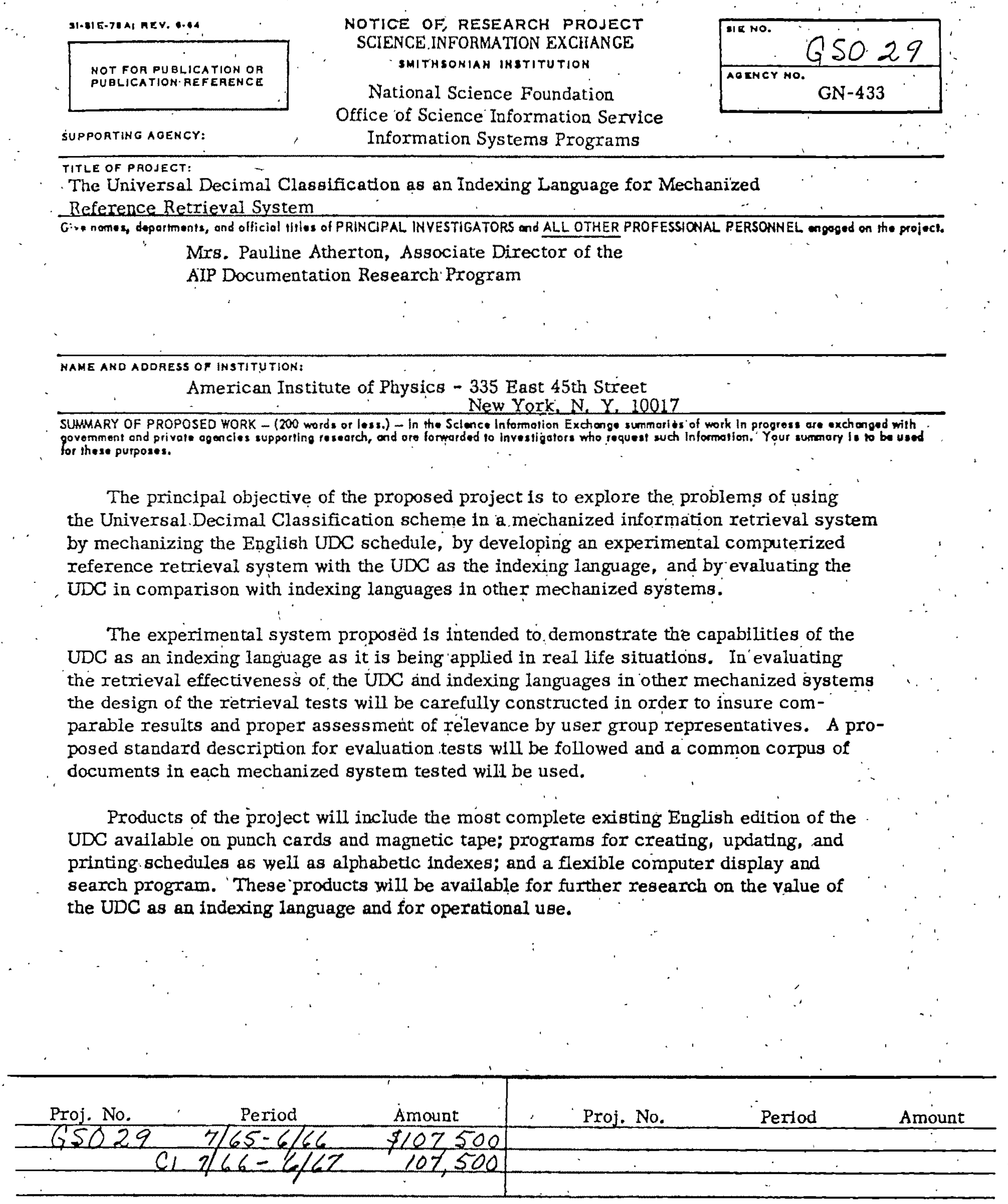
1. The supporting or funding agency.
2. The professional investigators.
3. The name and location of the researching institution.
4. The project title and a 200–300 word technical description of the research.
5. The period or beginning and end dates.
6. The annual funding level.

This information is put into the system on a Notice of Research Project (NRP) which is the unit hard copy records of SIE. Approximately 70,000 notices of current research were received by SIE in 1965. A sample NRP is shown in Fig. 1.

[1] Presented at the 1965 Congress of the International Federation of Documentation, Washington, D. C., October 15, 1965.

[2] Assistant Director, Operations, Science Information Exchange, Smithsonian Institution.

[3] Deputy Assistant Director, Operations, Science Information Exchange, Smithsonian Institution.



31-81E-78A; REV. 6-64

NOT FOR PUBLICATION OR PUBLICATION·REFERENCE

SUPPORTING AGENCY:

NOTICE OF RESEARCH PROJECT
SCIENCE INFORMATION EXCHANGE
SMITHSONIAN INSTITUTION

National Science Foundation
Office of Science Information Service
Information Systems Programs

SIE NO. GSO 29

AGENCY NO. GN-433

TITLE OF PROJECT:
The Universal Decimal Classification as an Indexing Language for Mechanized Reference Retrieval System

Give names, departments, and official titles of PRINCIPAL INVESTIGATORS and ALL OTHER PROFESSIONAL PERSONNEL engaged on the project.

Mrs. Pauline Atherton, Associate Director of the
AIP Documentation Research Program

NAME AND ADDRESS OF INSTITUTION:

American Institute of Physics - 335 East 45th Street
New York, N. Y. 10017

SUMMARY OF PROPOSED WORK – (200 words or less.) – In the Science Information Exchange summaries of work in progress are exchanged with government and private agencies supporting research, and are forwarded to investigators who request such information. Your summary is to be used for these purposes.

The principal objective of the proposed project is to explore the problems of using the Universal Decimal Classification scheme in a mechanized information retrieval system by mechanizing the English UDC schedule, by developing an experimental computerized reference retrieval system with the UDC as the indexing language, and by evaluating the UDC in comparison with indexing languages in other mechanized systems.

The experimental system proposed is intended to demonstrate the capabilities of the UDC as an indexing language as it is being applied in real life situations. In evaluating the retrieval effectiveness of the UDC and indexing languages in other mechanized systems the design of the retrieval tests will be carefully constructed in order to insure comparable results and proper assessment of relevance by user group representatives. A proposed standard description for evaluation tests will be followed and a common corpus of documents in each mechanized system tested will be used.

Products of the project will include the most complete existing English edition of the UDC available on punch cards and magnetic tape; programs for creating, updating, and printing schedules as well as alphabetic indexes; and a flexible computer display and search program. These products will be available for further research on the value of the UDC as an indexing language and for operational use.

Proj. No.	Period	Amount	Proj. No.	Period	Amount
GSO 29	7/65-6/66	$107 500			
C1	7/66 - 6/67	107,500			

FIG. 1. Notice of Research Project.

Notices are stored in several files:

1. Subject Files — A single NRP is filed under each subject index point to which the NRP has been indexed. This provides a source for rapid reference and/or retrieval for requests not requiring use of the computer.

2. Agency Files — A single NRP is filed by the agency supporting research and thus provides an internal reference source independent of the computer operation.

3. NRP Stacks — Multiple copies of each NRP are filed alphanumerically by accession number. This file is used to supply NRP's in hard copy when projects have

been identified by computer searches of the tape files or manual searches of the subject files. It is, per se, never used for search purposes.

Data from each NRP is also placed on appropriate magnetic tape files:

1. The Master File — An alphanumerical listing by SIE assigned accession number of a project record corresponding to a NRP. For a schematic format description see Fig. 2.
2. An Index File — A dictionary which contains codes and corresponding English captions for supporting agencies, locations, and subject index points. It is used for validating input codes and attaching captions to reports.
3. Contract No.–SIE No. Cross Reference File — Provides for manual crossover between agency identification and the accession number used by SIE.
4. Title File — An alphanumerical file of titles for all projects on the master file sequenced by SIE assigned accession number. It is used solely to attach titles to special reports.
5. Investigator File — Investigator names arranged alphabetically. It is used to respond to requests for all the projects on which particular investigators are working.
6. Pending Projects File — A list of proposed or pending projects. These records are addressable by accession number, investigator, and researching institution, but are not indexed to subject matter.

Except for searches on investigators' names, the master file is used for all mechanized searches. Depending upon the nature of the request, any or all of the other magnetic tape files may be brought into play. Some actual questions are shown in Fig. 3. In general, straight-forward subject, location, or agency searches need use only the master tape file because upon identification a list of accession numbers for the pertinent records is printed out. The NRP's are then pulled from the NRP stacks. If, however, a table or a compilation is to be generated in which English captions, titles, funds, or contract numbers are required, the other files are needed. These tasks almost always require programming and/or separate machine runs. These are the singly run jobs to be discussed more fully later.

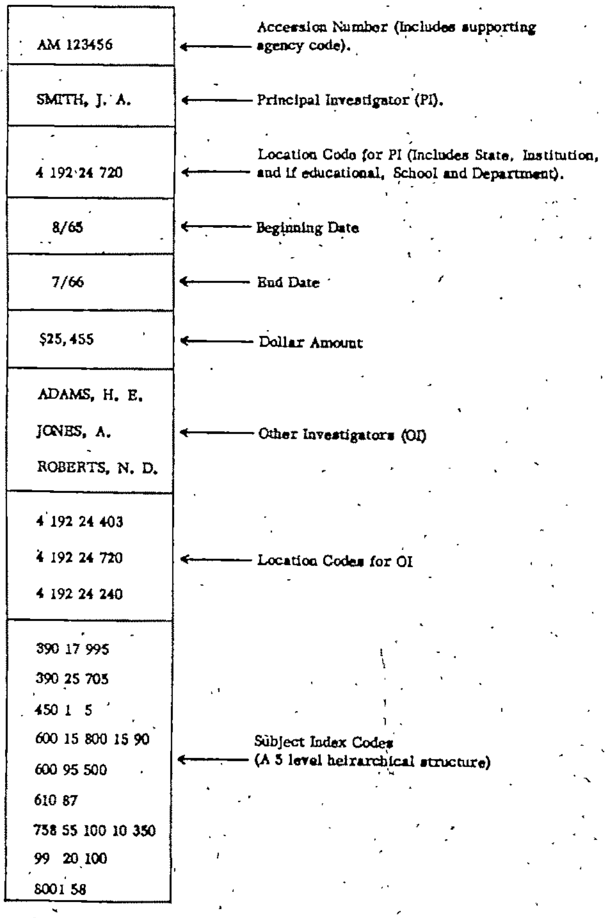


FIG. 2. Schematic magnetic tape record format.

SAMPLE QUESTIONS

Typical Subject Requests

1. Oxidations and autoxidation of long chain fatty acids.
   (SIE found 54 projects representing 19 different sources of support.)
2. Image covariance factor analysis.
   (SIE found 13 projects representing 7 different sources of support.)
3. Psychological stress in heart disease.
   (SIE found 36 projects representing 11 different sources of support.)
4. Electrochemistry of iron porphin complexes — e.g., Electron transfer rates, equilibrium, and formation rate constants.
   (SIE found 23 projects representing 8 different sources of support.)

Typical Administrative Requests

1. A tabulation of all federal grants (or contracts) to 50 select universities showing the total number of grants and funds to each department within the universities, prorated by Supporting Agency. Include subtotals for each graduate school.
2. An alphabetical listing of 1,400 neurosurgeons showing the research projects in which they are currently participating reflecting the title, dates and funds for each project.
3. An inventory of all current Public Health Service grants to Schools of Pharmacy.
4. An inventory of all U. S. supported research in Latin American countries arranged alphabetically by investigator within country; listing titles, dates, funds, supporting agencies and location of the project.

Typical Subject — Administrative Requests

1. Total funds spent on current research in the cardiovascular field in New York State.
2. A list of all current projects dealing with health related research in Foreign Countries which identifies the source of support.
3. All current studies on leukemia supported by Federal Government excluding USDA and NIH.

FIG. 3.

Computer Array

Until May 1965, SIE had a 16K, IBM 1401 computer. The hardware presently used is a 6 tape, 16K, IBM 1460 computer with a card-read punch, a 600 lines/minute printer, and a console typewriter attached.[4] The tape units are 729 model 5's operating at 800 bits/inch with a transfer rate of 60,000 characters/second. There is no disc capability.

Job Categorization

With very few exceptions, the tasks performed by the computer activity in an information center can be grouped for costing purposes into three types of jobs: batched, singly run, and file maintenance. Following is a discussion of how this grouping is performed at the Science Information Exchange.

1. *Batched Jobs.* These include all tasks which are grouped so that they can be performed concurrently during a single pass of the master files. Obviously, for economy and faster turnaround times, jobs are batched whenever possible. At SIE, almost all batched jobs involve multiple selection criteria and the superimposition of several boolean statements (matches). They are perhaps equivalent to conventional bibliographic compilations involving selection and organization by subjects, authors, or other search parameters. Fig. 4 shows a sample of an actual question asked of SIE and the corresponding "Request for a Computer Run" which was completed by a scientific analyst. It is a typical "batchable" question which involves four subject search terms (or parameters) and two matches (part B of the request).

Presently, programming and machine configuration limit the batch size to 150 search terms (i.e., subjects, locations, or supporting agencies). This may be one question with 150 parameters or 150 questions with one parameter or combinations thereof.

The cost of the individual job is computed by distributing the total batch processing time among each job in the batch in proportion to the number of subject terms each contains. The time is then multiplied by the hourly computer cost. It is, of course, recognized that this picture is oversimplified because there are other factors which affect total computer time per job. Analysis thus far, however, indicates that the number of search terms appears to be the dominant factor influencing job times.

2. *Singly Run Jobs.* This covers all tasks which use the files but which occupy the total data processing activity while being run. These jobs vary from very simple to extremely complex. Singly run jobs are performed as such only because either programming or machine limitations preclude batching. In such cases where part of a job is run in a batch and then selected material is further machined in order to arrange the information as required, it is counted as a singly run job with the batched time added to the time logged while it was being handled as a singly run job. Compilations involving several types of information (e. g., subjects, locations, and funds) almost always require special formating and are therefore handled as singly run jobs. Catalogs in which textual material or titles as well as subject, investigator, and location indexes are involved also must be handled as separate jobs and come within this category.

3. *Maintenance, Update and Research Tasks.* These are the computer runs which contribute to the quality and currency of the search files and the retrieval programs. Included are update runs and error correction passes on all files in current use as well as research and development of new file arrays or search programs. At SIE, the master file and all satellite files are updated every two weeks. This in itself is a major computer time commitment but is considered worth doing because stress is placed upon having an up-to-date master file which is as free from errors as is reasonably possible. Also, new and different demands require constant experimentation with new files and redesign of old ones for quicker, more expedient responses.

Computation of Unit Costs

If during a given period of time the total number of computer hours T is to be accounted for by the hours spent on file maintenance and research M, batched jobs B, and singly run jobs S, then obviously

$T=M+B+S$ (hours)

If D is the total dollar cost of the computer installation for that period, then

$d=\frac{D}{T}=$Average hourly cost (dollars per computer hour) and $C_M=Md$, $C_B=Bd$, and $C_S=Sd$ are the costs of running maintenance and research tasks, batched and singly run jobs respectively.

If during this time period there are n_1 batched jobs and n_2 singly run jobs, then

$$c_b=\frac{C_B}{n_1} \text{ and } c_s=\frac{C_s}{n_2}\left(\frac{\text{dollars}}{\text{job}}\right)$$

where c_b and c_s are unit costs for batched and singly run jobs respectively.

It should be noted that if the machine is under-utilized in this period in the sense of paid for time while the machine is idle, this is factored into the computation via D and thence to d, which represents the over-all cost. The burden is thus distributed over all the activities.

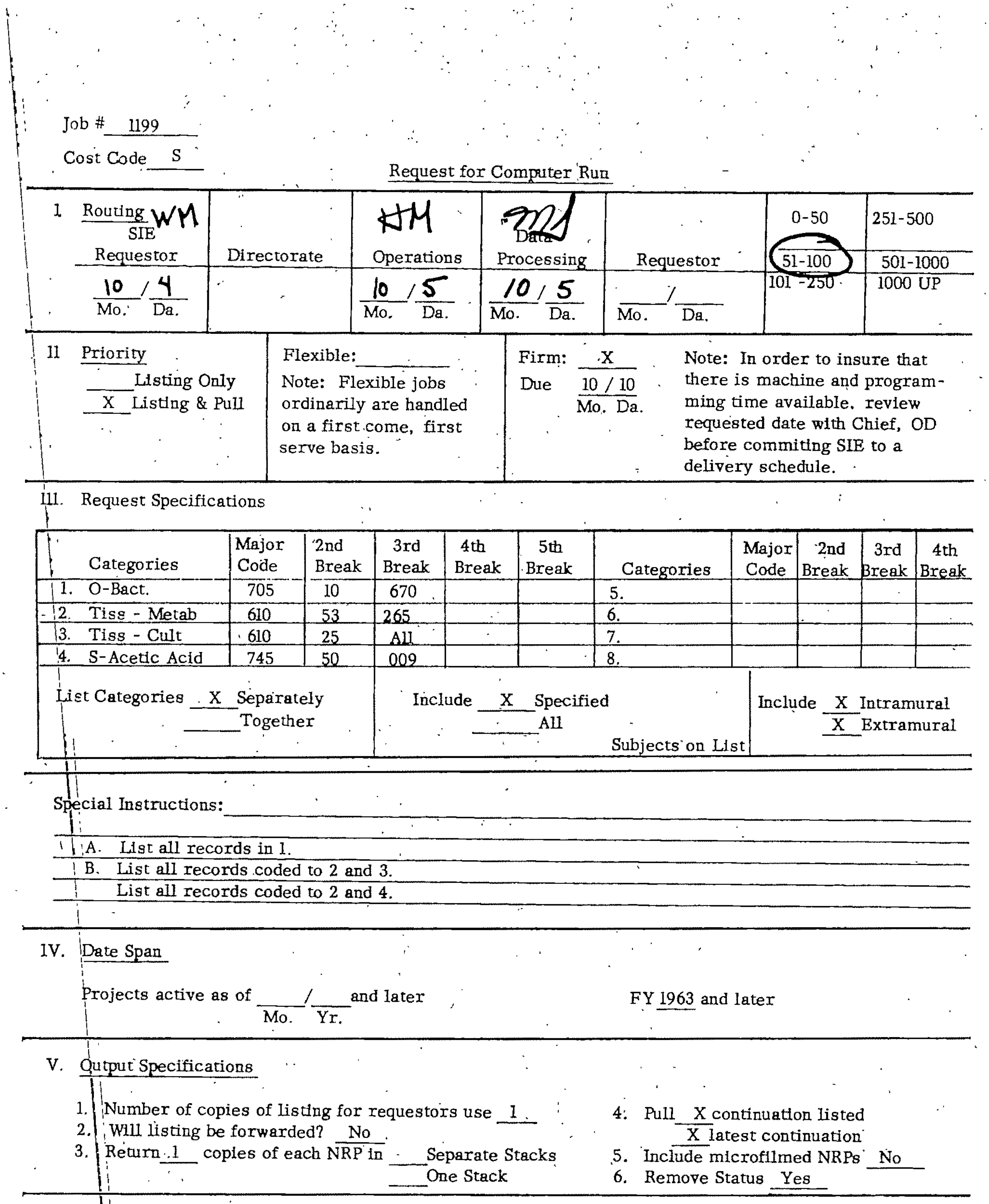
However, it is neither fair nor realistic to consider c_b and c_s as inclusive unit costs. File maintenance, update and research tasks are support activities which are done in order that batched and singly run tasks be performed efficiently on a current and accurate file. Therefore, it seems reasonable to lay off the file maintenance and search program costs against those jobs which use these files. Further, this lay off should be in proportion to the use of the files in which the maintenance investment is being made.

Accordingly C_M can be separated into two parts:

$$C_M=\left(\frac{B}{B+S}\right)C_M+\left(\frac{S}{B+S}\right)C_M$$

The adjusted unit cost for batched and singly run jobs

[4] In spring of 1966 this equipment will be replaced by IBM 360/30 equipment.



Job # 1199

Cost Code S

Request for Computer Run

1 Routing WM SIE Requestor	Directorate	HM Operations	Data Processing	Requestor	0-50	251-500
10 / 4 Mo. Da.		10 / 5 Mo. Da.	10 / 5 Mo. Da.	/ Mo. Da.	[✓ 51-100] 101 - 250	501-1000 1000 UP

II Priority

____ Listing Only
X Listing & Pull

Flexible: ____

Note: Flexible jobs ordinarily are handled on a first come, first serve basis.

Firm: X
Due 10 / 10
Mo. Da.

Note: In order to insure that there is machine and programming time available, review requested date with Chief, OD before commiting SIE to a delivery schedule.

III. Request Specifications

Categories	Major Code	2nd Break	3rd Break	4th Break	5th Break	Categories	Major Code	2nd Break	3rd Break	4th Break
1. O-Bact.	705	10	670			5.				
2. Tiss - Metab	610	53	265			6.				
3. Tiss - Cult	610	25	All			7.				
4. S-Acetic Acid	745	50	009			8.				

List Categories X Separately
____ Together

Include X Specified
____ All
Subjects on List

Include X Intramural
X Extramural

Special Instructions:

A. List all records in 1.
B. List all records coded to 2 and 3.
List all records coded to 2 and 4.

IV. Date Span

Projects active as of ___/___ and later
Mo. Yr.

FY 1963 and later

V. Output Specifications

1. Number of copies of listing for requestors use 1
2. Will listing be forwarded? No
3. Return 1 copies of each NRP in ___ Separate Stacks
___ One Stack
4. Pull X continuation listed
X latest continuation
5. Include microfilmed NRPs No
6. Remove Status Yes

FIG. 4. The Request for Computer Run for the question, "All research records dealing with pleuropneumonia-like organisms and acetate metabolism of tissue cells." It was prepared by an SIE scientific analyst. The "listing" is a computer-generated list of the pertinent NRP accession numbers, and the "pull" is the stack of NRP's. The latter will be forwarded, after screening, to the original questioner and the former retained for SIE record purposes. Other special requirements such as time span or scope of coverage are as indicated.

which include a prorated share of file maintenance costs becomes:

$$c_b' = \frac{C_B + \left(\frac{B}{B+S}\right) C_M}{n_1}$$

$$c_s' = \frac{C_S + \left(\frac{S}{B+S}\right) C_M}{n_2}$$

Upon examination of these formulas, it can be seen that unit costs are functions of several independent variables: total monthly computer usage (which determines d), ratio of the numbers of hours spent on maintenance, batched and singly run jobs, and the number and type of batched and singly run jobs. Obviously, the computation of unit costs under the best of circumstances is no easy matter. Indeed, prediction of unit job costs for short periods of time with a mix of jobs is necessarily inexact.

Also, the formulas show clearly what is intuitively expected, that as the ratio of maintenance hours to search hours increases and the number of jobs go down the unit job costs increase.

● Operating Experience

Fig. 5 shows the actual operating experience of the SIE for the full year 1964 and the first half of 1965. Considering the first portion of the table dealing with Direct Computer Costs (i.e., no maintenance costs are included), several aspects are worthy of note:

1. During the year 1964 the times occupied by maintenance, batched, and singly run jobs were 23%, 30.7%, and 46.3% respectively. The corresponding ratios during the first half of 1965 were 27%, 36.5%, and 36.5%. The trend in 1965 was towards a greater proportion of time spent on batched jobs and less on singly run jobs. The computer time spent on file maintenance tasks during 1964 and the first half of 1965 was roughly the same, about one quarter of the total computer load.
2. The average hourly computer cost was $45 in 1964 and $44 during the first half of 1965. The monthly variation, however, was considerable, going from a low of $37 to a high of $55. Obviously, greater computer use lowers the hourly rate. In May 1965, the IBM 1401 central processing unit was replaced with an IBM 1460 which rented for slightly more than the 1401 but processed all of the jobs slightly faster. The average hourly computer rate went up, but the unit batched job costs for May and June remained about the same. Apparently, the higher hourly rate was offset by the decrease in processing time.
3. In 1964, the average direct cost was $36 for batched jobs and $131 for singly run jobs. In the first half of 1965 the costs were $30 and $139. The authors feel the decrease in unit costs for batched jobs in the first part of 1965 is due primarily to better overall computer management and refinement of batched search programs. It is doubtful that these costs can be significantly further reduced without major system modifications.

A recomputation of unit costs with file maintenance cost laid off in proportion to the use of the file is shown on the right side of Fig. 5 under Total Computer Costs. The batched jobs unit costs jump from $36 to $48 for

					DIRECT COMPUTER COSTS (No Maintenance Costs)						TOTAL COMPUTER COSTS (Maintenance Costs Included)					
	Computer Use and Cost			Maint.	Batched Jobs			Singly Rum Jobs			Batched Jobs			Singly Run Jobs		
	Hrs.	Cost	$/Hr.	Hrs.	Hrs.	Jobs	$/Job	Hrs.	Jobs	$Jobs	Hrs.	Jobs	$/Job	Hrs.	Jobs	$/Job
Jan. '64	300	$12,500	$42	67	97	138	$29	136	33	$173	126	138	$38	175	33	$220
Feb.	221	11,700	53	55	50	88	30	115	21	290	67	88	41	153	21	386
Mar.	211	11,600	55	61	55	82	37	95	53	99	77	82	51	134	53	148
Apr.	229	11,900	52	71	74	71	53	84	49	88	107	71	79	122	49	127
May	319	13,200	42	72	105	96	46	141	34	172	136	96	59	182	34	222
June	307	13,000	43	58	118	154	33	132	27	207	145	154	40	162	27	254
July	285	12,300	43	66	81	96	36	139	31	193	105	96	47	181	31	250
Aug.	297	12,400	42	38	124	127	41	135	92	62	142	127	47	154	92	70
Sept.	289	12,350	43	67	85	102	36	136	54	108	111	102	47	178	54	142
Oct.	279	12,300	44	61	106	116	40	112	30	164	136	116	52	143	30	210
Nov.	287	12,350	43	47	66	95	30	174	61	123	79	95	36	208	61	147
Dec.	321	12,750	40	107	61	81	30	150	41	147	94	81	47	226	41	220
	3,345	$148,350	$45	770	1,022	1,246	$36	1,549	526	$131	1,325	1,246	$48	2,018	526	$170
Jan. '65	264	$12,150	$46	72	104	155	$31	87	36	$111	143	155	$43	120	36	$153
Feb.	289	12,450	43	97	104	131	34	88	24	157	157	131	52	132	24	236
Mar.	354	13,150	37	79	143	211	25	132	45	108	184	211	33	170	45	140
Apr.	324	12,800	40	92	115	144	32	118	17	274	160	144	44	165	17	382
May	243	12,850	53	78	98	157	33	67	31	115	144	157	50	99	31	170
June	254	13,100	52	48	67	134	26	139	47	152	83	134	32	171	47	187
	1,728	$76,500	$44	466	631	932	$30	631	200	$139	871	932	$41	857	200	$190

FIG. 5. SIE experience.

formation is displayed and the weight applied to each index term in retrieving documents will obviously influence the facility of communication, and consequently the utility of the index. However, even if these two factors were optimum, the utility of the index is still affected by the precision of the meaning of the indexing terms. Clearly a system with perfect display and optimum weighting factors is useless to someone who does not understand the meaning of the words that it uses.

Most information theorists seem to assume that meaning is perfectly precise; that is, they assume that concepts are lucidly embodied in documents, and that these concepts can be perceived and labeled with precision. To validate this assumption, they define or appoint a subject expert — someone whose duty it is to provide the standard indexing information. Invoking the subject expert is logically equivalent to the assumption that concepts in documents can be precisely labeled, at least by someone.

The concept of subject expert is unsatisfactory, both in practice and in theory. In practice, the available experts wish to restrict the area of their expertise to so narrow a speciality as to be nearly useless. Furthermore, no two subject experts agree, and neither is altogether comprehensible to the questioner. The indexer usually resorts to using himself as the subject expert, and to doing his best to explain his point of view to the questioner. Sometimes the point of view of the indexer is so foreign to the questioner that the latter will not attempt to use the index.[2]

As a practical expedient, the subject expert is not a useful addition to an information system. As a theoretical expedient, an average or abstraction of subject experts is not helpful either. We can do no better than assert that a subject expert is one who fulfills the assumption given above. Enough is known about the process of perception to make it most unlikely that a real person can fulfill that assumption.

In this paper, we adopt the point of view that a subject expert is neither necessary nor desirable, either as a part of an information system or as a standard of reference. We try to answer two questions: How precisely do ordinary people skilled in the subject perceive and label the concepts embodied in documents? Is this degree of precision such as to pose a difficult problem so severe as to make optimum weighting factors for index terms a minor improvement? In subsequent papers, I plan to investigate the improvement of communication between questioner and answerer as it is influenced by the content and layout of the index itself, always bearing in mind the precision with which the meaning of the indexing terms is understood by the questioners.

Hillman (2, 3) has shown that four factors enter into the definition of relevance: the question, the answer, a degree, and a corpus. The term "relevance" is ordinarily applied to define the relationship between two things, for instance, a question and answer. Many think that relevance either exists or does not exist. However, in many instances we must also admit to degrees of relevance. Out of the many possible answers to a question it is possible to say that some have a great deal of relevance, some less relevance, and others have little or no relevance. Hillman, in his paper, writes in terms of "concepts," suggesting that their interpretation depends on the field of knowledge (corpus) to which they are being applied. Hence, the field must be specified as a part of our definition of the relevance existing between two things. I should like to suggest that this expanded definition of relevance should also be applied to our consideration of words and their meanings. This is an intuitively agreeable course of action, since we are all aware of how a given word may have a number of meanings and of how the specific meaning applied to the word depends on the field of knowledge in which it is being used. I propose that meaning can be defined as the relevance of a word to the concept that it labels.[3] Therefore, if we are to specify the meaning of a word, we must also specify the four parameters suggested by Hillman's work: the word, the abstraction for which it is to stand, the degree of relevance that we wish to exist between the word and the abstraction, and the field of knowledge in which this word will be used.

How then can these ideas aid us to better understand the task of indexing? By assigning a descriptor[4] to a document, the indexer asserts that the descriptor has a high degree of relevance to the contents of the document; that is, he asserts that the meaning of the descriptor is strongly associated with a concept embodied in the document, and that it is appropriate for the subject area of the document. Let us assume that the indexers assign the descriptors in the order of the degree of relevance to the concepts, or that they assign all of the descriptors that they believe have a high degree of relevance. Then the consistency with which a given degree of relevance is associated with a given descriptor-concept pair will reflect the precision of the association strengths. Hence, consistency of indexing serves as a measure of the precision of meaning.[5]

Through measuring the consistency with which a term is applied to a concept, we are able to assess whether or not its meaning is understood with precision. By having a number of abstracts indexed by a number of people, it is possible to discover the consistency with which a given indexing term was used and, hence, how well the meaning of the term was understood.

[2] This is the situation of many organic chemists in regard to alphabetical indexes of chemical names.

[3] Harold Wooster suggests that meaning be defined as the degree of synonymy connecting a word and the concept it labels. However, I think it is clearer to use synonymy to name a relation between like things, e.g., one word has a degree of synonymy with another; and to use relevance to name a relation between unlike things, e.g., a word has a degree of relevance to a concept.

[4] In this paper, *descriptor* is used as a synonym for *index term* in accordance with popular usage, but at odds with its *formal definition.*

[5] For another approach to the measurement of the precision of meaning, see references 4 and 5.

• Part I. Indexing Consistency with Free Choice of Descriptors

In Part I of the experiment, 15 indexers were asked to choose descriptors for 50 abstracts which had been chosen at random from a single abstract journal. The indexers were to select any words or phrases that they considered appropriate. They were not supplied with instructions for making the choice, a word list, or the definition. This resulted in a list of descriptors that contained 1,050 different words and phrases that had been applied to the 50 abstracts. Plural and singular forms of the same word were counted as a single descriptor. The diversity of the responses made it impossible to use these data to arrive at any estimate of the precision with which the descriptors were used. However, several interesting conclusions could be drawn from these data. An average of 3.6 descriptors was assigned by each indexer to each abstract. However, let us remember that the depth of indexing is not the same as the number of index terms applied. Depth of indexing is measured by that proportion of concepts embodied in the document to which index terms are applied. If a document is about two concepts only, the deepest indexing can apply only two terms to it. Conversely, if a UDC number six digits long is applied to a document that embodies a dozen concepts, only one index term is applied — not the six terms implicit in the six digits — and the indexing is shallow. To determine the absolute depth of indexing for the documents used here, an absolute judgment of the number of concepts embodied in each document would be needed. Since we have avoided absolute judgment, we cannot make this calculation. To estimate this number of concepts by taking the consensus of the indexers, we needed the opinion of each indexer regarding the generic relations among the terms he applied — the degree to which each term implies or includes every other term. These opinions, rather difficult to form, were not determined. Consequently, we cannot estimate the depth of indexing and can only guess that it probably is not very deep.

Of the 1,050 descriptors chosen, only 48% of these were actual words or phrases that appeared in the abstract or the title. The indexers apparently found the language used by the author in the abstract inadequate for describing the work performed. As might be expected, a greater number of descriptors were applied to longer abstracts.

CHANCE OF RETRIEVAL

The descriptors assigned by the indexers to the subjects covered in an abstract *can be* considered to be the search program *they would* first formulate if they were attempting to find information on those subjects. The probability that the search devised by an average indexer would contain at least one descriptor in common with those chosen for a given abstract by all other indexers participating is given in Equation 1, where C_j is the chance of retrieval for abstract j, expressed in per cent, X_{ij} is the number of times the descriptor i was applied to the abstract j; n is the number of indexers; and m is the number of descriptors. This new measure of the chance for retrieval is necessary because existing measures of efficiency include a term which requires a judgment of the pertinence of the descriptor choice. No such judgment was made in dealing with these data.

Equation 1

$$C_j\ (\%) = \frac{\displaystyle\sum_{i=1}^{i=m} (X_{ij})^2 \qquad 2 \leq x \leq n}{\displaystyle\left[\sum_{i=1}^{i=m} X_{ij}\right]^2 \qquad 1 \leq x \leq n} \times 100$$

EQUATION 1. Chance of Retrieval for Abstract j, C_j.

If each indexer applies different descriptors to an abstract, the chance of retrieval is zero; if each indexer applies the same descriptors, the chance of retrieval is 100%. Nearly 62% of the descriptors were used only once; consequently, the chance of retrieval was low, varying from 1.5% to 12%, averaging 6.5%. This I consider to be a good estimate of the success of any initial search if it is made by a searcher familiar with, but not knowledgeable in, the special field of the search and if the searcher is not given indexing aids or the advice of subject experts. The chance of retrieval showed no correlation with the number of words in the title and in the text of the abstract. Of the approximately 1,050 descriptors used, 1%, or 10 words, accounted for 30% of the descriptor-abstract pairs and 2.2% of the descriptors accounted for half of the pairs.

This suggests that a much smaller number of descriptors would be almost as effective for describing these abstracts as the 1,050 actually used. The value of the chance of retrieval for each abstract was plotted versus the number of descriptors chosen for the abstract. The least-squares straight line showed a negative correlation, significant at 99% level. I conclude that, under the condition of free choice of descriptors, the greater the number of descriptors applied, the more difficult is the retrieval.

Some control of the indexing language is obviously the next step in our attempt to eliminate all the variables and measure only the precision of the meaning of the words — the precision with which indexers and questioners knowledgeable in the subject can perceive and label concepts embodied in documents. One possible mode of control would be to restrict the indexing language to that appearing in the documents. There are several objections to this procedure. First, it is not drastic enough. When one attempts to eliminate an undesired variable from a measurement, the best initial strategy is to give the undesired variable two extreme values. In this case, wishing to measure the precision of meaning independently of choice of language, we first give great freedom, then highly restrict choice of language. Free to use the language of the documents, the indexers can still use nearly half of the descriptors. As we shall see, a cut to 10% is not too drastic.

A second objection to the procedure of restricting the language of description to that of the documents arises from the particular document set used. Although many of the abstracts were written by the original authors, many were not. The language in the latter is not that of the author. Thus, we cannot restrict the indexer to the language used by the author. Furthermore, what we wish to estimate is the precision with which the indexer understands the meaning of the words, not the facility with which he picks out the author's words. In forcing the indexer to use any language but his own, we introduce an additional uncontrolled variable into our measurement. Finally, the author's language is not an optimum choice to reach conditions of maximum precision in meaning, because there is no reason to believe that authors use words more precisely than indexers do.

It seems best to begin with the list of freely chosen descriptors, for these represent the natural expression of the group of indexers, if not of each individual indexer. In restricting the list, we can make no "intelligent" choice, for to do so is to reintroduce the subject expert, whose influence we wish to avoid. The list should, then, be drastically reduced in size in an arbitrary way. The list was cut to one-tenth, using a random selection so arranged that terms used by many indexers are more likely to be retained. Chosen in this way, the final list resembles more closely the list used by each individual indexer than would be the case if the list were culled at random without consideration of the frequency of use of the terms.

To examine the effect of reducing the number of descriptors on the chances of retrieval, a small set of descriptors was chosen from the total list of 1,050. Descriptors were chosen for this set by first giving a weight to each descriptor — a weight representing, not its probable utility in retrieval, but the frequency with which it was applied by the indexers. Numbers from 1 to 2,180 were assigned to each descriptor each time it was applied, so that a given descriptor that was applied to 30 abstracts might receive the numbers 1–30 or 70–99, whereas a descriptor that was applied only once would be given only a single number. Numbers were then taken from a random number table (6) and matched against the numbers assigned to the descriptors until a set of 100 descriptors was chosen. This set of descriptors was then compared with our original descriptor-abstract data to find how many times each descriptor in the set of 100 had been applied to the 50 abstracts by the original 15 indexers. We then substituted this new frequency data in Equation 1 to find how restricting the number of descriptors had affected the chance of retrieval of each abstract. This chance of retrieval is the probability that the first search program would succeed if only the descriptors in the selected list of 100 were used. This chance varied from 0 to 100%, averaging 36%. When the selected vocabulary was used to calculate the chance of retrieval, the correlation with the number of descriptors applied rose to a small positive value, indicating no correlation. That is, the disadvantage of assigning many descriptors to a document, without control of choice, was overcome by restricting the choice of descriptors used in searching (Table 1).

• Part II. Indexing Consistency with Restricted Choice of Descriptors

In Part II of the experiment, the same 50 abstracts, rearranged in a random fashion, were sent to 9 indexers, all of whom were among the original 15 indexers. These indexers were now supplied with the randomly chosen list of 100 descriptors and requested to apply these to the abstracts. For several of the abstracts, no adequate descriptors appeared in the list; consequently, the subjects were asked whether or not they considered the descriptors they had chosen for each abstract to constitute an adequate description. This opinion showed no correlation with the average number of descriptors chosen, and none with the chance of retrieval.

Restricting the choice of descriptors results in a marked increase in the consistency (7, 8, 9) with which they are applied. In Part I, no descriptor was applied to any abstract by all of the indexers, whereas in Part II, 19 of the descriptors were applied by all of the indexers. However, of these 19 only 6 were applied with perfect consistency; that is, they were either applied 9 times or not at all. The other 13 descriptors were applied consistently to certain abstracts by all indexers but inconsistently to other abstracts. Of the 100 descriptors, 15 describe con-

TABLE 1. Relation of the number of descriptors used for retrieval to retrieval chance and to the correlation of retrieval chance with the number of descriptors assigned to an abstract.

Number of Descriptors Used	Chance of Retrieval, % Min.	Max.	Aver.	Correlation Z Value	Significant
Part I 1000	1.5	12	6.5	—0.51	Yes
100	0	100	36	+0.04	No
Part II 100	9.8	56	22	—0.73	Yes
50	0	67	33	—0.52	Yes
25	0	100	48	—0.35	Yes
12	0	100	45	—0.13	No
6	0	100	23	*+0.27*	*No*
Part III 45	8.4	15	12	+0.03	No
22	14	18	16	+0.17	No
11	19	91	37	+0.27	No
6	37	44	41	—0.19	No

cepts that were unknown only a few years ago. Five of these new words, or 33% of them, were among the 19 most precise descriptors, while only 16% of the older words were used precisely. It would be interesting to know if new concepts are understood more precisely than older ones as suggested by these data.

Twelve of the 100 descriptors on the list were used in 34% of the descriptor-abstract pairs. Of these often-used descriptors, only 1 (or 75%) was new. The other 11 descriptors, or 13% of the total, were older words. This suggests that descriptors for older concepts tend to be used more frequently.

The chance of retrieval was calculated for each abstract by means of Equation 1. It varied from 9.8% to 56%, averaging 22%, a substantial improvement over Part I. The correlation of the chance of retrieval with the number of descriptors assigned to the abstract was again strong and negative. Smaller sets of descriptors were chosen from the list of 100 by using the weighting method described in Part I. Sets of 50, 25, 12, and 6 descriptors were chosen. Each set was then used to calculate the chance of retrieval for each abstract with the results shown in Table 1, and graphically in Fig. 1. Again, restricting the number of descriptors excludes the possibility of retrieving some of the abstracts, i.e., the chance of retrieval is 0, but it increases the average chance of retrieval for the sets of 50 and 25 descriptors (dashed curve). However, with the sets of 12 and 6 descriptors, more and more abstracts are irretrievable and the average chance of retrieval diminishes. The correlation between chance of retrieval for a particular abstract and the number of descriptors applied to the abstract is strongly negative for the largest three sets of descriptors (100, 50, 25) and insignificant for the smallest two sets.

• Part III. Indexing Consistency with Many Indexers

For the third part of the experiment 21 abstracts were selected. Each abstract was one about which there had been substantial disagreement among the indexers as to whether descriptors assigned from the list of 100 adequately described the abstract or not. From these 21, the abstracts were chosen that required the use of "new" descriptors; i.e., terms that have been added to the technical vocabulary during the last few years. This eliminated all but 8, and from these 5 were selected at random. The descriptor set made up for these abstracts consisted of all the descriptors that had been assigned to these abstracts by indexers in Part II. This required 28 descriptors. Seventeen additional descriptors were then chosen from the Part II set at random by using the weighting procedure already described to compile a list of 45 descriptors. The abstracts and descriptors were sent to several hundred indexers with backgrounds similar to those of the indexers used in the first two parts of this experiment. Each indexer was asked to assign descriptors from the list of 45 to each of the 5 abstracts and to indicate whether or not they considered the descriptors chosen adequately described the abstracts.

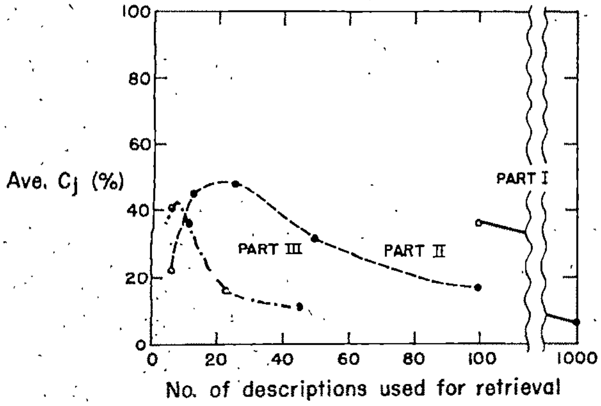


Fig. 1. Relation of the number of descriptors used in retrieval with the average chance of retrieval.

Three hundred and twenty-three indexers responded. There were several differences between the results of Part II and those of Part III and the conclusions that could be drawn from them. The chance of retrieval in Part III was lower than that in Part II; the larger number of indexers had a greater difference of opinion, lowering the chance of retrieval. The correlation of chance of retrieval with the number of descriptors assigned per abstract is insignificant — probably a consequence of the drastic reduction in the number of abstracts. The 12 descriptors most often used accounted for 59.5% of the total number, or nearly twice the fraction of Part II. Of these, 6 were new and 4 showed good precision. The difference in usage of new words and the tendency to use new words less precisely than old words fail to appear in Part III, probably because the particular abstracts were chosen to include frequent use of new words.

No descriptor was used in Part III with perfect precision. The best four descriptors (lead monoxide, modulation transfer function, phase modulation, nuclear reactors) were used by 96%, 93%, 88%, and 85% of the indexers, respectively, for one of the 5 abstracts, and none of the indexers applied any of these terms to more than a single abstract.

The typical graph shown in Fig. 2 shows how "lead monoxide" was handled in each of the three parts of this experiment. As we have already noted, the descriptor was applied in Part III to one of the 5 abstracts by 96% of the indexers and was not applied to any of the other 4 abstracts by any of the 323 indexers. This is shown by the graph labeled "Part III" in Fig. 2. In Part II in which 9 indexers were asked to index 50 abstracts using a list of 100 terms, "lead monoxide" was applied to one abstract by 67% of the indexers and to a second by 78%. None of the indexers applied this term to any of the other 48 abstracts. In Part I in which 15 indexers were asked to index the same 50 abstracts, but without the aid of a descriptor list, the term was applied by 53% of the

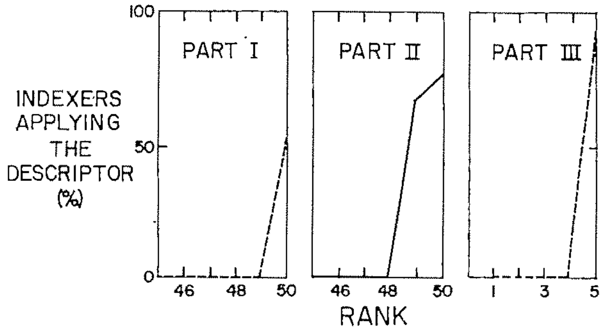


FIG. 2. Graphs illustrating the precision of meaning of "lead monoxide."

indexers to a single abstract and it was not applied to any other abstract by any of the indexers.

At this point, it is interesting to observe the behavior of the indexers with respect to some of the other terms. Fig. 3 shows how the term "data processing" was applied. Note the perfect precision in the graph of Part II with which this descriptor was applied. However, in Part I the term was applied by far fewer indexers when they were required to supply the term rather than to assign the term from a list. In Part III, for some reason, the term was applied somewhat less frequently to the one abstract (i.e., less frequently than in Part II), but more frequently to two of the other abstracts. Since the population of indexers is so small in Part II, the results in Part III probably express more accurately the precision with which this term is applied or understood.

An objection might be raised that, although the meaning of a word may be perfectly understood by an indexer, he might fail to apply the word because he considers it useless for retrieval of the particular abstract. This objection is specious: the distinction is too fine to influence the results of this experiment. That is, for the purposes of this experiment, the following statement was taken as a tautology: The word that stands for a concept is useful in retrieving an abstract treating that concept. The terms "lead monoxide" and "data processing" can be used to illustrate this point of view. If a given term is applied to a specific abstract by a large number of indexers, it is fair to say that those who do not apply the term do not fully understand its meaning. Hence, we can say that the consistency with which a term is applied by a large number of indexers is a good measure of how well that term is actually understood.

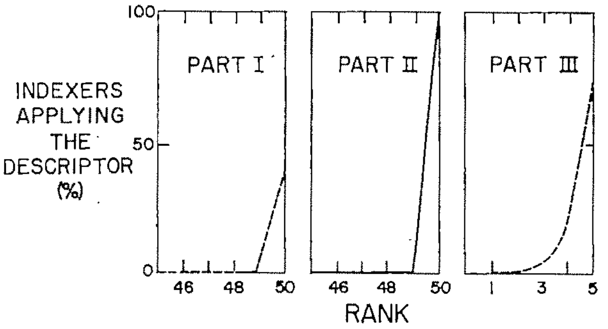


FIG. 3. Graphs illustrating the precision of meaning of "data processing."

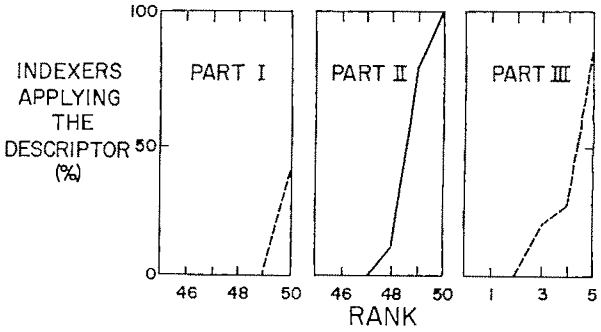


FIG. 4. Graphs illustrating the precision of meaning of "electrophotographic plates."

Next, an attempt was made to discover whether or not a group of indexers who used one term with precision used other terms with equal precision. To do this, the term "electrophotographic plates" was selected because it had been used throughout the experiment with average precision. Fig. 4 shows how this term was applied to a single abstract by 88% of the indexers. We now reject the 12% of the indexer population that failed to apply the term to abstract 5, and an additional 34% minority who applied it to two other abstracts. By examining how the remaining 54% (174 indexers) applied other terms, we can gain some understanding of whether or not a population that is consistent in its use of a given term is also consistent in its use of other terms. Fig. 5 shows how the term "conductivity" was handled in the three phases of the experiment. The right-hand graph shows how the selected group of 174 indexers applied "conductivity" to the 5 abstracts used in Part III. We note that the selected group, who had been consistent in the application of the term "electrophotographic plates," is not consistent in its application of the term "conductivity." If this same procedure is applied to other pairs of terms, we always find this behavior, i.e., that a given term is applied consistently by a group of indexers is no guarantee that the group will apply some other term with equal consistency.

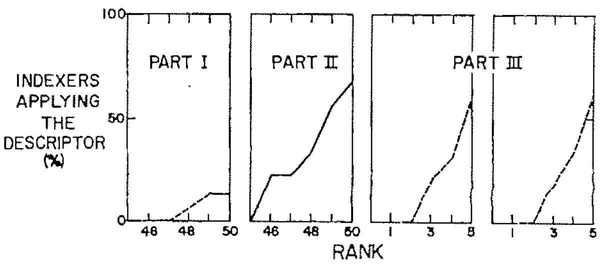


FIG. 5. Graphs illustrating the precision of meaning of "conductivity." The graph on the extreme right shows the results in Part III, with all assignments deleted for indexers who disagree on "electrophotographic plates."

Summary

In Part I of the experiment, 15 indexers chose descriptors for 50 abstracts with no restrictions on the choice of descriptors or on the language with which they were expressed. Many descriptors were chosen, although only a moderate depth of indexing (3.6 descriptors per abstract) resulted. Only about half of the descriptors matched words or phrases of the abstracts. A new measure of retrieval efficiency was needed because no judgment of correctness of the indexing was made. The chance that at least one descriptor applied by one indexer would match the descriptors applied by all the other indexers, the chance of retrieval, was calculated. For the 50 abstracts the average chance of retrieval is 6.5%. The average chance of retrieval rose to 36% (nearly six times as high) when calculated using data from a selected list of 100 terms. In Part I, even though the field of knowledge (corpus) and the concepts embodied in the abstracts were controlled, the consistency of indexing was low. The variety of expressions used for an identical or closely related concept frustrated the attempt to relate precision of meaning to indexing consistency.

In Part II, 9 of the same indexers applied descriptors to the same 50 abstracts, using only the 100 selected terms. The consistency of application increased markedly, and 6 of the terms were used with perfect precision. The chance of retrieval averaged 22%. When an even smaller selected list of descriptors (25) was used, the chance of retrieval rose to a maximum of 48%.

Most descriptors were used imprecisely. Since meaning is defined in terms of the relevance of a word to the concept it labels, descriptors will be imprecisely applied if any one of four factors (the field of knowledge, the concepts, the words, or the degree of relevance) is not controlled. All indexers must have a common understanding of the concepts used in a given corpus of knowledge and the words to be associated with these concepts. They must also have a common understanding of the degree of association (relevance) that exists between a word and the concept for which it stands. Analyses of the results show that three of these factors (field of knowledge, the concepts, and the words) were adequately controlled under the conditions of this experiment. However, because only 9 indexers were used, the degree of relevance cannot be measured with a high degree of confidence. Hence, a larger group of indexers was used in Part III in order to demonstrate more fully the connection between precision of meaning and consistency of indexing.

In Part III of the experiment, 323 indexers applied 45 selected descriptors to 5 abstracts. The consistency of application decreased and the chance of retrieval averaged 16%. Using fewer descriptors from this list raised the chance of retrieval. Graphs, in which the fraction of indexers applying a given descriptor to the abstract was plotted versus the abstracts and ranked so that the highest ranking abstract had the greatest fraction of indexers applying the particular descriptor, illustrate the precision with which the descriptor is used. None of the descriptors in Part III was used with perfect precision. Furthermore, it was shown that precision in the use of one term did not imply that the same indexer would use other terms with precision.

In the experiment, the field of knowledge was constant and well defined, the terms were kept constant, and the concepts embodied in the abstracts were kept constant. Nonetheless, perfect consistency was not obtained. Obviously, a fourth variable is present. I conclude that meaning can be defined as the relevance of a word to the concept it labels, and that the degree of relevance, as it is understood by the various indexers, is the most important variable accounting for these results.

It is clear that the meaning of words is not understood precisely, even when the words labeling scientific concepts are used by competent scientists in their own field of knowledge.[6] Any aid (thesaurus) or procedure(multiple indexing, personal contact) that helps to increase the precision of meaning among a particular group can be expected to increase the effectiveness with which they exchange information. To make an excellent index means must be found that will allow good communication by means of words which are inherently imprecise.

References

1. Martyn, J. 1964. Literature Searching by Research Scientists. Aslib Res. Dept. 3 pp.
2. Hillman, D. J. On Concept—Formation and Relevance. *Proc. Am. Doc. Inst. Meeting, October 5–8. 1964.* Vol. 1: 23–29.
3. Hillman, D. J. 1964. The Notion of Relevance. *Am. Doc.*, 15 (1): 26–34.
4. Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. 1957. *The Measurement of Meaning.* Univ. of Ill. Press, Urbana, Ill.
5. Forster, K. I., Triandis, H. C., and Osgood, C. E. 1964. *An Analysis of the Method of Triads in Research on the Measurement of Meaning.* Tech. Report No. 17, June. AD 603 944. Dept. of Psychology, Univ. of Ill., Urbana, Ill.
6. RAND Corporation. 1955. *A Million Random Digits with 100,000 Normal Deviates.* Free Press, Glencoe, Ill.
7. Jacoby, J., and Slamecka, V. 1962. *Indexer Consistency Under Minimal Conditions.* November. AD 288 087.
8. Slamecka, V. *Final Report, Indexing Aids.* 1963. January. AD 294 859.
9. Borko, H. 1961. *A Research Plan for Evaluating the Effectiveness of Various Indexing Systems.* July. AD 278 624.
10. Chapanis, A. 1965. Color Names for Color Space. *Am. Scientist*, September, pp. 327–346.

[6] A similar suggestion has been made by Alphonse Chapanis in reference 10.

Brief Communication

A Matrix for Evaluating Information System Operation[1]

There are so many sources of information available and so many types of inquiries that it is difficult to talk explicitly about information exchange in a research field. The simple technique presented here is the approach we used to examine information exchange relating to civil defense research. It may prove useful to others in examining information systems.

Information Users

The matrix shown as Fig. 1 illustrates present information transfer patterns. For illustration, five users of civil defense information are listed down the left side of the figure.

1. Office of Civil Defense (OCD)
2. State and Local Civil Defense
3. Research Contractors
4. Delegate Government Agencies (Those agencies with specific civil defense responsibilities as delegated by Executive Order of the President.)
5. Public and Other Users

Forms of Information

Nine forms of information are listed beside each of the five users. These forms show that information is obtained in many ways.

A user can know what information exists by reading *Review Articles*, by regularly scanning *Accession Lists*, or by requesting *Bibliographies* on a particular subject.

The user may also be led to information by *Discussions* with individuals knowledgeable in the field, particularly those individuals with research *Work in Progress*. *News Releases* may also notify users of available information.

The information itself may be contained adequately in *Abstracts* or it may be necessary to read the *Reports and Books* themselves. The information also may be obtained by attendance at *Formal Meetings* where papers are presented.

The nine terms in italics are referred to in Fig. 1 as the forms of information available to each of the five users. These forms correspond to the rows in the information exchange matrix.

The Sources of Information

Nineteen sources or types of sources of information are listed across the top of Fig. 1. Two of these sources are hypothetical; i.e., they don't exist at present, but may be useful. These sources form the columns of the information exchange matrix.

1. Research Directorate—OCD
2. Other Personnel—OCD
3. Publications—OCD
4. Information Center—OCD (Hypothetical Services)
5. Depository Libraries—OCD (Hypothetical Services)
6. Contractors—OCD
7. Army Library—Pentagon
8. Defense Documentation Center
9. Delegate Agencies (Including OEP)
10. Government Printing Office
11. Clearinghouse for Federal Scientific and Technical Information
12. Library of Congress
13. Atomic Energy Commission
14. NIH, NASA, NSF, etc.
15. Science Information Exchange
16. National Academy of Science (NRC)
17. Centers for Analysis of Scientific and Technical Information
18. Professional Societies and Journals
19. News Media

Explanation of the Matrix in Fig. 1

The present information transfer patterns are evaluated in the matrix presentation of Fig. 1.

Our evaluation of information exchange is placed in the rectangle formed by the intersection of a row and a column. For instance, row one described the places a user in the Office of Civil Defense presently searches for a report copy: chiefly, the OCD Research Directorate, the Army Library, or the Defense Documentation Center. Occasionally this user requests the document from a Contractor, a Delegate Agency, the Government Printing Office, the Federal Clearinghouse, the Atomic Energy Commission, or the National Institutes of Health.

In other words, when a solid black rectangle falls at the intersection of a user row and a source column, it means that the user normally looks to that source for information. In 6the illustration, a user in the Office of Civil Defense when looking for reports normally looks to the Army Library (among others) for copies. Therefore, a black rectangle fills the space at the intersection of row one, (Office of Civil Defense—Reports & Books) and column seven (Army Library—Pentagon).

When a black dot is placed at an intersection, it means that some limited exchange occurs. The limitation may occur because the source service is limited or because the users make inadequate use of the available service.

When no black mark is placed at an intersection, little or no information exchange now occurs between the particular user and source combination.

Explanation of the Matrix in Fig. 2

Fig. 2 illustrates the exchange which might occur under ideal conditions. It recognizes that some services are not being used because they are not and will not be appropriate.

[1] This article is based on research sponsored by the Office of Civil Defense, Department of the Army, under contract OCD-PS-64-56. At the time this research was conducted Mr. Jenkins and Mr. Herzog were operations research analysts with the Research Triangle Institute, Durham, North Carolina.

THE USERS OF CIVIL DEFENSE INFORMATION AND THE FORMS OF THIS INFORMATION ↓	THE SOURCES OF INFORMATION USEFUL TO CIVIL DEFENSE RESEARCH →	OCD-RESEARCH DIRECTORATE	OTHER PERSONNEL	PUBLICATIONS	INFORMATION CENTER (PROP.)	DEPOSITORY LIBRARIES (PROP.)	CONTRACTORS	ARMY LIBRARY - PENTAGON	DEFENSE DOCUMENTATION CENTER	DELEGATE AGENCIES (INCLUDING OEP)	GOVERNMENT PRINTING OFFICE	CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION	LIBRARY OF CONGRESS	ATOMIC ENERGY COMMISSION	NIH, NASA, NSF, ETC.	SCIENCE INFORMATION EXCHANGE	NATIONAL ACADEMY OF SCIENCE (NRC)	CENTERS FOR ANALYSIS OF SCIENTIFIC AND TECHNICAL INFORMATION	PROFESSIONAL SOCIETIES & JOURNALS	NEWS MEDIA
OFFICE OF CIVIL DEFENSE	REPORTS & BOOKS	■					●	■	■	●	●	●		●	●					
	REVIEW ARTICLES																			
	ABSTRACTS						●		■			●			●					
	ACCESSION LISTS							■	■		●	●		●	●					
	BIBLIOGRAPHIES						●		●			●	●		●			●		
	NEWS RELEASES			●						●				●	●					●
	FORMAL MEETINGS	■	■				■			●				■			■			
	DISCUSSIONS	■	■				■			●				●			●			
	WORK IN PROGRESS	●	●				■			●					●					
STATE AND LOCAL CIVIL DEFENSE	REPORTS & BOOKS	●						●			■	●								
	REVIEW ARTICLES																			
	ABSTRACTS											●								
	ACCESSION LISTS							●			■	●								
	BIBLIOGRAPHIES											●	●							
	NEWS RELEASES		●							●				●	●					●
	FORMAL MEETINGS		■							●									●	
	DISCUSSIONS	●	●																	
	WORK IN PROGRESS																			
RESEARCH CONTRACTORS	REPORTS & BOOKS	●					■	■	■	●	●	●		●	●					
	REVIEW ARTICLES																			
	ABSTRACTS							●	■			●								
	ACCESSION LISTS							■	■		●	●		●	●					
	BIBLIOGRAPHIES								●			●	●							
	NEWS RELEASES									●				●	●					●
	FORMAL MEETINGS	■	●				■							●					●	
	DISCUSSIONS	■	■				●			●										
	WORK IN PROGRESS																			
DELEGATE GOVERNMENT AGENCIES	REPORTS & BOOKS	●						●	●	●	■	■		●	●					
	REVIEW ARTICLES																			
	ABSTRACTS								●			■								
	ACCESSION LISTS							●	●		■	■		●	●					
	BIBLIOGRAPHIES											●	●							
	NEWS RELEASES									●				●	■					■
	FORMAL MEETINGS		●												●		●		●	
	DISCUSSIONS	●	●				●			●				●	●			●		
	WORK IN PROGRESS																			
PUBLIC AND OTHER USERS	REPORTS & BOOKS										■	■								
	REVIEW ARTICLES																			
	ABSTRACTS											■								
	ACCESSION LISTS										■	■								
	BIBLIOGRAPHIES												●							
	NEWS RELEASES									●				●	●					●
	FORMAL MEETINGS																		●	
	DISCUSSIONS	●	●							●										
	WORK IN PROGRESS																			

FIG. 1. Present transfer patterns, civil defense information system.

Clear rectangle = Little or no information is presently provided to the user by the corresponding source.
Black dot = There is limited to fair exchange of information at present.
Black rectangle = There is satisfactory or good exchange of information at present.

THE USERS OF CIVIL DEFENSE INFORMATION AND THE FORMS OF THIS INFORMATION ↓ / THE SOURCES OF INFORMATION USEFUL TO CIVIL DEFENSE RESEARCH →		OCD-RESEARCH DIRECTORATE	OTHER PERSONNEL	PUBLICATIONS	INFORMATION CENTER (PROP.)	DEPOSITORY LIBRARIES (PROP.)	CONTRACTORS	ARMY LIBRARY - PENTAGON	DEFENSE DOCUMENTATION CENTER	DELEGATE AGENCIES (INCLUDING OEP)	GOVERNMENT PRINTING OFFICE	CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION	LIBRARY OF CONGRESS	ATOMIC ENERGY COMMISSION	NIH, NASA, NSF, ETC.	SCIENCE INFORMATION EXCHANGE	NATIONAL ACADEMY OF SCIENCE (NRC)	CENTERS FOR ANALYSIS OF SCIENTIFIC AND TECHNICAL INFORMATION	PROFESSIONAL SOCIETIES & JOURNALS	NEWS MEDIA
OFFICE OF CIVIL DEFENSE	REPORTS & BOOKS	■		●	■		●	■	■	■	●	●		■	●			●	●	
	REVIEW ARTICLES	■	■	■	■					●	●			●			■		●	
	ABSTRACTS			■	■		●	■	■	●		●			●					
	ACCESSION LISTS				■			■	■	●	●	●		●	●					
	BIBLIOGRAPHIES				■		●	■	■	●		●	■	●	●			■		
	NEWS RELEASES			●						●				●	●					■
	FORMAL MEETINGS	■	■				■			■				■			■		■	
	DISCUSSIONS	■	■		●		■			■				■	●		●	●		
	WORK IN PROGRESS	■	■	■	●		■	●		■				●	●	■		●		
STATE AND LOCAL CIVIL DEFENSE	REPORTS & BOOKS	●		●	■	■		●		●	■	■		●	●			●	●	
	REVIEW ARTICLES	●		●	■					●	■			●			●		■	
	ABSTRACTS			●	■							■			●				●	
	ACCESSION LISTS				■			●			■	■		●	●					
	BIBLIOGRAPHIES				■	■		●				■	■					●		
	NEWS RELEASES		●							■				■	●					●
	FORMAL MEETINGS		■				●			■									■	
	DISCUSSIONS	●	●		●		●			●				●				●		
	WORK IN PROGRESS		●		●	●		●		●						■			●	●
RESEARCH CONTRACTORS	REPORTS & BOOKS	■		●	■	●	■	■	■	■	●	●		●	●			●	●	
	REVIEW ARTICLES	■		■	■		●			■	■			●			■		■	
	ABSTRACTS			●	■			●	■	●		●							●	
	ACCESSION LISTS			■	■			■	■	●	●	●		●	●					
	BIBLIOGRAPHIES				■	●		■	■			●	●					■		
	NEWS RELEASES									●				●	●					●
	FORMAL MEETINGS	■	●				■			●				■			●		■	
	DISCUSSIONS	■	■		●		■			■				●				●		
	WORK IN PROGRESS	●	●	■	■			■		●				●		■			●	●
DELEGATE GOVERNMENT AGENCIES	REPORTS & BOOKS	●					●	■	■	●	■	■		●	●			●	●	
	REVIEW ARTICLES				●					●	■						●		●	
	ABSTRACTS			■					■			■							●	
	ACCESSION LISTS							■	■	●	■	■		●	●					
	BIBLIOGRAPHIES							■	■			■	■					●		
	NEWS RELEASES			●						●				●	■					■
	FORMAL MEETINGS		●											●	●		●		■	
	DISCUSSIONS	●	●		●		●			●				●	●			●		
	WORK IN PROGRESS			●				■		●						■			●	●
PUBLIC AND OTHER USERS	REPORTS & BOOKS			●	●	■					■	■							■	
	REVIEW ARTICLES										■						●		■	
	ABSTRACTS											■							■	
	ACCESSION LISTS			●							■	■								
	BIBLIOGRAPHIES				●	■						■	■					●		
	NEWS RELEASES									■				●	●					■
	FORMAL MEETINGS																		■	
	DISCUSSIONS	●	●		■	●				●								●		
	WORK IN PROGRESS			●	■											●			●	●

Fig. 2. Potential transfer patterns, civil defense information system.

Clear rectangle	=	There is little or no potential exchange of information.
Black dot	=	There is limited to fair potential for information exchange in the future.
Black rectangle	=	There is good potential for information exchange in the future.

Others are not being used, but should; e.g., Science Information Exchange.

If Fig. 1 is printed in color on a transparent overlay and placed on top of Fig. 2 the opportunities for improving the information exchange patterns become immediately obvious. The black rectangles and dots of the potential exchange that show through the overlay are points where the potential is not being met. The overlay provides a simple method of visualizing the overall information exchange pattern and for identifying existing gaps in the information transfer process.

Conclusions

Many scientific and technical information services are available from various government and private agencies. Consequently, numerous formal and informal channels of information exchange exist which should be considered before adding new information services in a research field.

These existing channels are not easy to visualize. We found the matrix approach useful in studying the overall exchange of information, in identifying new systems which might be useful, and in identifying existing services that can be used more efficiently.

J. Edward Jenkins[2] and
William T. Herzog[3]
Research Triangle Institute
Durham, North Carolina

[2] Presently with Department of Electrical Engineering, University of Edinburgh, Edinburgh, Scotland.

[3] Presently Assistant Professor, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.

Letters to the Editor

Dear Sir:

Writing of Venn diagrams, Mr. Sharp (1) asserts, "It is, unfortunately, impossible to use simple diagrams for more than 4 terms."

He may well be correct about their use or their utility. Nevertheless, their construction is possible for any number of variables, using simply connected areas that are at least conceptually simple also.

Many methods have been described and reinvented since 1881, most of them using comb-like polygons. This indeed was suggested by Venn himself.

The 5-term diagram results from placing a U-shaped polygon on the well-known 4-term set of rectangles. This method can be extended indefinitely. See, for instance, Anderson and Cleaver (2), and also Martin Gardner's book (3). The latter is fully illustrated and has a good bibliography.

I take this opportunity to remind those who need to be reminded that Euler diagrams are not Venn diagrams.

References

1. Sharp, J. R., 1966. The SLIC index. *Amer. Doc.*, **17** (1): 41–44, Jan.
2. Anderson, D. E., and Cleaver, F. L. 1965. Venn-type diagrams for arguments of N terms. *J. Symbolic Logic*, **30** (2): 113–118, June.
3. Gardner, M. 1958. *Logic Machines and Diagrams.* Chapter 2. McGraw-Hill, New York.

Robert Fairthorne
Herner and Company
Washington, D. C.

Dear Sir:

In the July 1965 issue of *American Documentation* (Volume 16, No. 13), on page 237, in the paragraph "International," there are some strange statements about ICSU A.B. and F.I.D., on which I would like to draw your attention, since you are among the officers of the "American Documentation Institute."

(1) We are not considering to extend our activities to include Nuclear Energy and Geology.

(2) As you may see, there is some bad muddle about the Study on Abstracting Periodicals in Physics.

All what is said in the second paragraph "The International Federation . . . . Conference of Biological Editors" are not F.I.D. recommendations, but recommendations of the UNESCO final Technical Wording Group on Scientific Documentation which met in Paris in March 1964. The report of this Working Group is document UNESCO/NS/Doc. TWG 2.

1. *Co-ordination of the recommendations of the three Working Parties on Scientific Documentation — Ways and Means of implementing the above recommendations.*

The recommendations of the three Working Parties in scientific publications (Philadelphia, September 1963), automatic documentation (Moscow, November 1963) and scientific translation and terminology (Rome, January 1964) were examined and generally supported and approved. In addition, the following comments and suggestions were agreed:

Ad hoc Sub-committee to study and report on methods of primary scientific publication. It was stressed that it would be advisable to have on this Sub-committee representatives of editors, scientific documentalists and librarians, specialists in computation and users (scientists and engineers). Names of possible members of this Sub-Committee were proposed but it was finally decided that the members of the Working Group should send to the Secretary a short list of names of individuals and organizations that might be represented on the Sub-committee. It was stressed also that the terms of reference should include not only primary publications but also their relation to secondary publications.

It was considered that the study proposed by the ICSU Abstracting Board on abstracting periodicals in physics — "Bulletin Signalétique", "Physikalische Berichte", "Referativny Zhurnal" — would provide a good source of information for the Sub-committee. It was suggested that Unesco should therefore assist the project, which will complement a similar survey that is being carried out in the USA on Physics Abstracts.

I think it would be necessary that *American Documentation* publishes an errata about these points. We would like that the statement on the ICSU A.B. be as follows:

The ICSU A.B. continued its general activities to improve the dissemination and quality of scientific information. The service of exchange of proof copies or advance copies for Member Services of the Board (the main abstracting periodicals all over the world covering Physics, Chemistry and Biology) has been reorganized and enlarged. The ICSU A.B. is also doing some precise studies, among which may be quoted:

— a detailed study of the main primary periodicals in Physics, Chemistry and Biology.

— a complete statistical study of the 1964 issues of Physics Abstracts, Physikalische Berichte and Bulletin Signalétique (Physics sections).

Results of these studies will be published in the course of 1966.

J. Poyen
Conseil International des
Unions Scientifiques
Paris

Book Reviews

The reviews in this issue were written as one of the regular assignments in the Modern Information Systems course at the Columbia School of Library Service. They are of books important in the field which have not previously been reviewed in American Documentation, *and are included here primarily as reviews — but also as some kind of reflection of education in the field. Student reviews are solicited from other institutions as well. Interested prospective reviewers, practitioners, and students are again urged to write to the Reviews Editor, indicating their special areas of interest.*

4/66–1R **Textbook on Mechanized Information Retrieval.** 1962. Allen Kent. Interscience, New York. 268 pp.

In his preface Kent delineates five purposes to be served by his textbook: (1) to serve as a textbook for fifth year graduate study in library school; (2) to serve the administrator, scientist, and practicing librarian in acquiring some basic understanding of a field that has progressed so fast that it may have eluded their ken; (3) to serve the developer of a retrieval system, whether for individual use or for large-scale exploitation, who cannot obtain unbiased advice as to choice of procedures and equipment; (4) to serve as a first guide for those who wish to compare their retrieval system to others that may be used; (5) to serve commercial interests who, in the long run, will benefit from an educated clientele.

This is a very large order and one that has been only partially fulfilled. Basically, this is a textbook for beginners. It can help students, administrators, scientists, librarians, and those interested in developing a retrieval system, *provided* they are beginners in the field of information retrieval. In the fourth purpose as outlined, Kent himself qualifies the use of his book as a first guide for those who wish to compare their systems with others. To say that the book will benefit commercial interests through educating readers to the uses of equipment is stretching a point, but may be true if, indeed, more librarians are inspired to apply some of the knowledge gained.

The book is divided into two sections. Section I contains eight chapters of text illustrations. Section II is made up of supplementary reading lists, classroom exercises, field trips, suggestions for the use of audiovisual material, and a sample examination.

After a chapter of Introduction (corresponding to Bourne's "The Nature of the Problem") describing the information problem and what the book is about, Kent devotes a chapter to "Physical Tools." This chapter provides a step-by-step discussion of the unit operations in machine literature searching, and it is from this unit operation approach that Kent writes his book. A brief list of the headings in this chapter will show how simply and clearly the material has been handled. Unit operations include: analysis; control of terminology; recording results of analysis in a searchable medium; storage of source documents, extracts, abstracts, bibliographic references; question analysis and development of search strategy; conducting the search; delivery of results. Each of these headings is further subdivided and discussed.

With this material there are excellent illustrations of the various tools. Using this chapter as a basis, Kent then devotes Chapter 3 to more detailed discussion of the "Principles of Analysis" (including indexing) and Chapters 4 and 5 to the "Principles of Searching" and the "Manipulation of Searching Devices." Kent's exposition in Chapter 3 of indexing by means of the two techniques, "word indexing" and "controlled indexing," seems more meaningful than Bourne's chapter on "Classification and Indexing." This may only prove that Kent's book is for the novice, for Bourne's chapter goes into far more detail and gives more examples of indexing systems. An example of Kent's elementary approach can be found in his step-by-step detailed discussion of a method of producing a concordance. This type of explanation is not found in Bourne, who assumes the reader has the basic knowledge and, if he does not, Bourne gives him bibliographic citations, e.g., Wisby, R. "Concordance Making by Electronic Computer: Some Experiences with the 'Wiener Genesis'," *Modern Language Review,* **57,** (2): 161–172 (April 1962).

In Chapters 4 and 5 of Kent, the strategies of searching, and how the various non-conventional systems are used to satisfy these strategies, are especially illuminating. These logical concepts are not discussed by Bourne. As noted earlier, it is the step-by-step explanations in Kent that are helpful to one who is ignorant of the field. For example, to obtain the logical product of logical sums $(A + B) \cdot (C + D)$ with marginal punched cards, Kent writes: "The cards are sorted for aspect A; those not selected are sorted for aspect B. Those selected after each sort are sorted for the presence of aspect C; those selected are the first fraction of the response. Those not selected are sorted for aspect D; those selected are the second fraction of the response."

Chapter 6, "Words, Language, Meaning and Retrieval Systems," is weaker and somewhat repetitious of the information on indexing already presented. Chapter 7, "Codes and Notations," cannot be compared with Bourne's similar chapter in erudition, thoroughness, and scope. It is good for basic simple principles. It differentiates in simple language between "superimposed and nonsuperimposed recording of notations," "numeric versus alphabetic notations," "fixed field versus free field," etc.

"System Design Criteria," Chapter 8, offers factors, such as one's objectives, functions, performance requirements, and environmental variables, to consider when designing a system. It is a very general guide offering no comparison of what various systems can do nor how much they might cost.

Part II, devoted to "Exercises," seems quite elementary when compared with the paper required for some introductory courses. One which might be cited is to have the class sit in a circle and whisper one to another to note how information is changed during transmission. A suggestion in Part II that might be worthwhile, however, was to show a film on machine literature searching. This might be a supplement to a trip to IBM, as it would be geared to library-type operations. Films suggested were those produced by the General Electric Company; Smith, Kline and French; or the Armed Services Technical Information Agency.

The bibliography in Kent's book is most parochial — 85% of the references are to articles or books written by Kent or by his associates. It could not be used for deeper study of the subject as could Bourne's excellent bibliography.

Both an author and subject index are provided. This *Textbook on Mechanized Information Retrieval* could profitably be read before Bourne as an introduction to the field of information handling. It offers simple definitions and step-by-step procedures. Illustrations, charts, and samples are good. However, the unit operation approach gives no picture of systems as a whole, such as is found in Bourne, nor does it adequately consider system applications or evaluations. Kent is especially weak in information regarding computers. For a more scholarly and complete reference work with excellent bibliography one must use Bourne. For anyone developing a retrieval system to rely solely on Kent would be foolhardy. A library school course completely dependent on Kent without the addition of Bourne is not exploring information systems in depth and is not using the basic tool in the field.

AUDREY RUBIN

4/66-2R **Towards Information Retrieval.** 1961. Robert A. Fairthorne. Butterworths, London. 211 pp.

Towards Information Retrieval is a collection of papers written over a thirteen-year period. While no attempt has been made to add textual material which would unify them, the papers do form a composite picture of some of the theoretical aspects of information retrieval, bringing to the reader various facets of a common theme.

Robert A. Fairthorne is a noted figure in a documentation, although perhaps better known in England than in the United States. The earliest paper presented in this collection dates from 1947 indicating his long concern with the field. Some thirty-five years of the author's career were spent at the Royal Aircraft Establishment, where he was consultant at the time of this publication. During the first twenty years of his career he applied his mathematical training to a wide variety of technical problems. During the next fifteen years he was also much involved with the organization of the library at the Royal Aircraft Establishment and the practical and theoretical problems involved became his major concern. He is now with Herner and Company in the United States.

In his article "Identifying Key Contributions in Information Science" (*Am Doc* **15**: 289–295, Oct. 1964), Carlos A. Cuadra singles out this volume as a major text in the field. Fairthorne's name also appears in Cuadra's table of frequently cited authors, derived from a count of entries in bibliographies in the field in the same study. The Cuadra study shows the article "Basic Postulates and Common Syntax" which appears in this volume as being cited by five major textbooks in the field.

However, the book cannot be considered as a text in the field in the sense of providing a survey of the entire field and fundamental information on its key aspects. It does, nevertheless, represent the best (according to the reviews surveyed) and probably the most frequently cited of Fairthorne's works. The theme which ties the paper together is probably best described in Fairthorne's own words taken from the preface:

> For some millenia librarians have had to deal with texts as carriers of concepts, and with texts as heavy objects with marks on. They have evolved efficient techniques and principles to cope with these aspects severally. Rarely have they discussed texts in both capacities at once.
>
> The selection of papers published here explores activities in which indefinite neglect of either aspect, the conceptual or the mechanical, will lead to practical and theoretical disaster. They centre on the recovery of records according to their subject matter.

The articles explore various areas of documentation, analyze and criticize existing systems, and seek new insights for blending the conceptual with the manipulative. Throughout the papers, Fairthorne's intent appears to be that of raising problems and drawing attention to them rather than offering solutions. In an introductory section entitled "Comments," Lea M. Bohnert states that the best introduction to the field and to the author's general approach is the paper "The Pattern of Retrieval" originally published in *American Documentation* in 1956 and reprinted here. Fairthorne indicates the nature of his concerns when he writes, "A deep question of great theoretical and practical importance is how far can we go in documentation, as in computing, by using ritual in place of understanding?" On notation: "The bridge between the concepts and physics of retrieval is notation, or systems of marking the texts." "They [librarians] have given little attention and have had little need to give attention to the mechanical consequence of notation considered as instruction for retrieving rather than recognizing documents." Fairthorne spends some time in discussing the classification of tasks in information work, especially those which, in his words, may be "delegated" to the machine. "Fortunately," says Bohnert, "Fairthorne belongs to the economic breed that considers it efficient to have human machines perform the unusual and variegated types of work." Fairthorne does not appear to expect classification to solve the problems of retrieval as seems currently fashionable. In fact, he does not seem to expect much from classification at all, in spite of his several writings on the subject.

Another of his major concerns is that of cost. He states that theory can be used to produce a fair estimate of costs when we study "all the links in the operational chain." But: "The theory can give only the least cost of clerical operations. Evidently the greatest cost depends only on what the author of the system can get other people to put up with. In practice, the limit seems to have been reached by the time the entries needed for retrieval exceed those in the documents to be retrieved." He believes that models of document retrieval systems should be used for experimental study before more money is spent on expensive varieties of retrieval machinery.

On the whole the volume is not easy reading. Much of it is theoretical and requires slow deliberate concentration for comprehension, and even then some is elusive. Fairthorne works through what he has to say with precision. His mathematical interests and ability are evident in the many diagrams and formulas. To the mathematically untrained the volume appears rather frightening by its not infrequent complicated passages.

Yet Fairthorne cannot be criticized for being deliberately obscure, or unnecessarily complex. The writing is straightforward and lucid. He apparently attempts to write with great clarity — so much so that he often achieves a disarming simplicity in his statements. His tendency to reduce complex notions to ordinary terminology is often evident — "marking" and "parking" as the two physical methods for organizing information for the retrieval process. He is often amusing or witty. When he is critical, his criticism is often biting, as in the opening of his article on "Delegation of Classification."

This volume is a most valuable contribution to the literature. That librarians have failed to appreciate Fairthorne can, according to Vickery, be attributed to a number of factors: "By and large, librarians are concerned to emphasize the intellectual content of their work, and display a marked psychological resistance to a description of part of it as 'clerical' and capable of performance by automation. It almost seems that they spurn labour-saving devices, despite their constant complaint of overwork. They have a fear of automata to overcome. They should ponder Fairthorne's words 'automatism is merely remote control in time.'" Vickery adds that Fairthorne has never actually participated in building a retrieval system. "In short, he is a theorist, and suffers the usual fate of lack of understanding by 'practical men.'"

The index is by Calvin Mooers.

While much of the material is not recent, most of Fairthorne's questions are as valid today as they were when he first raised them. This is true mainly because the author's concern is with basic theory, and not with descriptions of current practice.

MARGARET LINN

4/66-3R **Indexing Theory, Indexing Methods and Search Devices.** 1964. Frederick Jonker. Scarecrow, New York. 124 pp.

Frederick Jonker's chief purpose in writing this book was to give a full exposition of a "generalized theory of indexing" which he had begun to develop a few years earlier. The expression "generalized theory" may be understood as referring to the process of describing a group or series of events in words sufficiently general to encompass all aspects of those events, and sufficiently specific that the description is recognizable as being uniquely of those events. Some groups of events lend themselves quite readily to such treatment by exhibiting many characteristics in common. An author then says that he has formulated a theory, because he has noticed the common characteristics. Needless to say, once a theory has been formulated, it is very easy to see subsequent events as operating within its framework. Indeed, sometimes it is almost impossible not to see them that way. Moreover, when events are described in the terminology of the theory, they seem subtly to change in character to fit it.

Jonker points out that the only valid criterion for designing an information retrieval system is cost: "how to deliver a specified quantity, quality and speed of service at the lowest possible cost." Since the initial indexing, and not the entering of data or the search, is by far the greatest cost factor, an analysis and understanding of this part of the I. R. structure is the prime necessity. However, since he is attempting to provide the common precepts by which individual systems can be judged for their suitability to a particular I. R. problem, the author has formulated his theory to cover all aspects of the systems.

Jonker begins the development of his theory by defining mechanized I. R. as "search by coincidence of terms." He proves this by demonstrating that all the logical relationships among indexing terms which a system may be required to provide may be reduced to readout functions, to coincidence-of-terms search, or to a combination of the two. He does not, however, limit his theory to a description of these activities. He points out that most systems in actual use are combinations of hierarchical or classified grouping with term coordination, and therefore attempts to encompass both.

The two basic factors in any index are the kind of terminology used, and the ways in which the words are made to relate to each other and to indicate relationships among the concepts embodied in the information store.

For the first of these, Jonker postulates a "terminological continuum" which he conceives as a direct function of the development of knowledge. He represents it schematically as a straight line proceeding from left to right. It is his contention that the language of a field of knowledge develops from longer to shorter terms. When a new concept is born and recognized, words are taken from several older concepts to describe it. He considers this the left end of the continuum. As the new concept becomes accepted and widely used, and in turn forms the basis for further developments in the field, new and unique words are used to describe it. Sometimes two or more older-concept words are simply combined to form one; sometimes they are hyphenated into a single inseparable expression; in other cases, a new word is coined. Since this is a natural evolutionary process, it cannot be depended upon to happen consistently, or at a particular rate of speed. It does not eliminate ambiguities caused by synonyms, homonyms, and shades of accepted meaning when the same word is used in different but related fields. It does not obviate the problem that different people in referring to the same concept will use words from different stages of development of the vocabulary. For greatest precision, therefore, an indexing system should, in principle, assign a unique word or code indication to every unique concept. This is the extreme right end of the continuum.

Such accuracy can be achieved only at great cost. In practice, the theory has two lessons for the system designer. He must be aware of the level of the vocabulary development (within the field with which the system deals) of the users of the system. He must also understand the language of the body of literature to which he is providing access. His job is to create a bridge between the language of the user and that of the system and, then, through really effective indexing, between the system and the literature. The first span may consist simply of a list of the index terms used; it may be in the form of a thesaurus; or it may be a translation mechanism built right into the machine.

The author puts forward cogent arguments against the possibility of a universal indexing vocabulary, applicable to all fields and all users. He claims that there is no standard criterion upon which to base such a language. In some cases, the use of something may be the best way to describe it. In others, for other people, the structure of that same thing may be more important.

Jonker postulates a "connective continuum" to describe the historical development of ways of showing relationships among concepts indexed. At one end is the classification system where a term is placed with others to which it bears a hierarchical relationship. This produces very long index terms, since each one carries with it all its relatives. At the opposite end is the keyword technique. Here, the only thing that can be discovered is what other items of information are stored in the same document or what other documents bear the same information. This produces the shortest index terms. Between the two ends is the subject heading list that overlaps both extremes by frequently giving some hierarchical indication, while also giving several subject headings for a particular item of information. The greatest potential for "indexing depth" (defined by the author as the number of criteria by which an item of information may be indexed) is at the short-term end of the continuum. However, since an item of information is entered at this end only on the hierarchical level on which it appears in the literature being indexed, it will be lost in a search by a word applying to a higher or lower level. Searches must be made on various levels if generic information is sought. On the other hand, the short-term end can handle ideas at all levels of their development, since as many terms as deemed necessary can be used to describe them, with no need to fit them into a preconceived pattern. The system designed at the short-term end of the continuum is, therefore, inexpensive (relatively) to feed, but may incur great expense in search time or coordination mechanisms at the output. A classified system is more expensive to feed, and may lose new ideas by erroneously placing them in hierarchies in which they are later found not to belong, but should be the simplest and cheapest at the output.

Integrating the two continua, any I. R. system might be viewed as a point on a two-dimensional plane. The horizontal might be considered the indexing type moving from classification to keyword, and the vertical the terminological type, moving up from lay to professional language (long to shorter terms). The decision must always be made at which point along each of these lines a particular system will operate. Lines drawn from these points, and perpendicular to the axes, will intersect at a point which may be said to define the system.

Developing his theory further, Jonker goes on to analyze the mechanization of I. R. systems. At the present time, he feels, the most important time- and labor-saving functions of mechanization lie in faciliating term correlation. Other operations, such as automatic encoding, printout, etc., are simply added benefits in more complex systems.

For students of information systems, perhaps the most valuable section of this book is the chapter entitled "Primary design consideration" (pp. 90–114). Here, the author gives a clear discussion of existing commercially available systems from the point of view of their suitability to particular I. R. needs. Working from his theory, he considers the efficiency with which they accomplish term correlation with respect to the way they handle three basic operations:

store organization (document or term grouping)

matching (simultaneous or sequential)

access to the store (single or multiple)

Using this gauge, eight basic types of systems are possible, in the form of different combinations of these operations. The author gives examples of systems embodying each of the combinations. He gives excellent diagrams which demonstrate the principles by which they work, and which are far more valuable than the photographs in, for example, Bourne's *Methods of Information Handling*. There is also a list of sources of supply, but with none of the valuable cost information given by Bourne.

Jonker seems somewhat overly impressed with the theoretical approach. He makes the following statement about the machine "art": "In developing his design, the designer proceeds from the most fundamental considerations available to him to considerations which are usually of a less abstract nature, and from there to design details" (p. 85). This is a theoretically sound approach, but, in practice, if the process takes place, it must often be somewhere below the conscious level. Nevertheless, this book has much valuable material for the student, and the "general theory" may be at least *a* way to analyze the systems available when a potential consumer must decide which one suits him best.

EDITH WARD

CALL FOR PAPERS

The 29th Annual Meeting of the American Documentation Institute will be held in Santa Monica, California, on October 3 through 7, at the Miramar Hotel.

Theme of the meeting is "Progress in Information Science and Technology." We will welcome papers reporting on original research, significant trends, and new concepts, techniques, and applications of information science and technology. You are cordially invited to help create an interesting, varied, and informative technical program.

Accepted papers will appear in the Proceedings, to be available one month before the conference, and will be the focus of individual Author Forums.

An award will be given by ADI for the three best papers submitted for the meeting. To insure maximum visibility for exceptional work, these papers will also be read by the authors at a special plenary session.

Contributed papers may be up to 2,500 words in length and may include illustrations not to exceed two printed pages. Papers should be accompanied by a 100- to 125-word abstract. Five copies of paper and abstract should be submitted by May 15, 1966. Contributors will be notified regarding acceptance of papers by August 1, 1966.

Send To: Dr. Carlos A. Cuadra
Technical Program Chairman, 1966 ADI Meeting
System Development Corporation
2500 Colorado Ave.
Santa Monica, Calif.

PROGRAM OUTLINE

TUTORIAL SESSIONS – October 3, 1966

Information Systems Design – R.M. Hayes – UCLA
Information Center Operations – A. Kent – University of Pittsburgh
Usage of Information – S. Herner – Herner and Co.
Evaluation of Hardware and Software – To be announced
Language Data Processing – H.P. Edmundson – SDC
Development of a Theory – D. Hillman – Lehigh University

STUDENT PROGRAM – October 3, 1966

Special Session – Student Papers
Panel Discussion – Student Chapter Activities
Cocktail Hour

J. Harvey – Chairman Student Membership Committee

PROGRESS REVIEW SESSIONS – October 4-7, 1966

Professional Aspects of Information Science and Technology – R.S. Taylor – Lehigh University
Information Needs and Uses – H. Menzel – New York University
Content Analysis, Specification and Control for Document Retrieval Systems – P. Baxendale – IBM
File Organization and Search Techniques – D. Climenson – U.S. Government
Man-Machine Communication – R.M. Davis – Dept. of Defense
Evaluation of Indexing Systems – C.P. Bourne – Programming Services, Incorporated
Automated Language Processing – R.F. Simmons – SDC
New Hardware Developments – M.E. Stevens – National Bureau of Standards
Information System Applications – J. Baruch – Bolt, Beranek and Newman
Library Automation – D.V. Black – University of California, Santa Cruz
Information Centers and Services – G.S. Simpson – Batelle Memorial Institute
National Information Issues and Trends – J. Sherrod – Atomic Energy Commission

SPECIAL FEATURES

Author Forums
Discussion Groups
Prize Papers
Special Libraries Association Session
Placement Service
Exhibits
Information Theater
Chapter Officers Workshop

Special Interest Groups
Proceedings
Award of Merit
Exhibitor Presentations
Tours
Evening in Disneyland
Buffet Luau

Address additional inquiries c/o
Technical Program Chairman

American Documentation

JULY 1966

Vol. 17, No. 3

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

AS WE MAY THINK

by VANNEVAR BUSH

As Director of the Office of Scientific Research and Development, DR. VANNEVAR BUSH has coördinated the activities of some six thousand leading American scientists in the application of science to warfare. In this significant article he holds up an incentive for scientists when the fighting has ceased. He urges that men of science should then turn to the massive task of making more accessible our bewildering store of knowledge. For years inventions have extended man's physical powers rather than the powers of his mind. Trip hammers that multiply the fists, microscopes that sharpen the eye, and engines of destruction and detection are new results, but not the end results, of modern science. Now, says Dr. Bush, instruments are at hand which, if properly developed, will give man access to and command over the inherited knowledge of the ages. The perfection of these pacific instruments should be the first objective of our scientists as they emerge from their war work. Like Emerson's famous address of 1837 on "The American Scholar," this paper by Dr. Bush calls for a new relationship between thinking man and the sum of our knowledge. — THE EDITOR

THIS has not been a scientist's war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. Now, for many, this appears to be approaching an end. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indecision, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who edge of his own biological processes so that he has had a progressive freedom from disease and an increased span of life. They are illuminating the interactions of his physiological and psychological functions, giving the promise of an improved mental health.

Science has provided the swiftest communication between individuals; it has provided a record of ideas and has enabled man to manipulate and to make extracts from that record so that knowledge evolves and endures throughout the life of a race rather than that of an individual.

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers — conclusions which he cannot find time to grasp, much less to remember, as they appear.

AMERICAN DOCUMENTATION

INSTRUCTIONS TO AUTHORS

American Documentation is a publication of the American Documentation Institute. It is a scholarly journal in the various fields in documentation and serves as a forum for discussion and experimentation. Papers already published or in press elsewhere are not acceptable. For each proposed contribution, one original and two copies (in English only) should be mailed to Mr. Arthur W. Elias, Editor, *American Documentation,* Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106. The manuscript should be mailed *flat* in a suitable-sized envelope. Graphic materials should be submitted with suitable cardboard backing.

Types of Manuscripts: Three types of contributions are considered for publication: full-length articles, brief communications of 1,000 words or less, and letters to the editor. Letters and brief communications can generally be published sooner than full-length manuscripts. Books, monographs, and reports are accepted for critical review. Two copies should be addressed to the Review Editor, Dr. T. Hines, 54 North Drive, East Brunswick, New Jersey.

Processing: Acknowledgment will be made of receipt of all manuscripts. *American Documentation* employs a reviewing procedure in which all mansucripts are sent to two referees for comment. When both referees have replied, copies of their comments are sent to authors with the Editor's decision as to acceptability. The refereeing procedure requires about 30 days. Authors receive galley proofs with a five-day allowance for corrections. Standard proofreading marks should be employed. Reprint order forms are forwarded with galleys.

Format: All contributions should be typewritten on white bond paper on one side only, leaving about 1.25 inches (or 3 cm) of space around all margins of standard, letter-size (8.5 × 11 inch) paper. Double spacing must be used throughout, including the title page, tables, legends, and references. The first page of the manuscript should carry both the first and last names of all authors, the institutions or organizations with which the authors are affiliated, and notation as to which author should receive the galleys for proofreading. All succeeding pages should carry the last name of the first author in the upper right-hand corner (0.5 inch from the top) and the number of the page.

Style: In general, style should follow the forms given in the Style Manual for Biological Journals (SMBJ), published for the Conference of Biological Editors by the American Institute of Biological Sciences (1964).

Title: The title should be as brief, specific, and descriptive as possible. Vague and unrevealing titles may delay publication.

Abstract: An informative abstract of 200 words or less must be included, typed with double spacing on a separate sheet. This abstract should present the scope of the work, methods, results, and conclusions.

Acknowledgments: Financial support may be listed as a footnote to the title. Credit for materials and technical assistance or advice may be cited in a section headed "Acknowledgments," which should appear at the end of the text. General use of footnotes in the text should be avoided.

Graphic materials: *American Documentation* requires finished artwork. Follow the style in current issues for layout and type faces in tables and figures. A table or figure should be constructed so as to be completely intelligible without further reference to the text. Lengthy tabulations of essentially similar data should be avoided.

Figures should be lettered in black India ink. Charts drawn in India ink should be so executed throughout, with no typewritten material included. Letters and numbers appearing in figures should be distinct and large enough so that no character will be less than 2 mm high after reduction. A line 0.4 mm wide reproduces satisfactorily when reduced by one-half. Graphs, charts, and photographs should be given consecutive figure numbers as they will appear in the text; however, figure numbers and legends should not appear as part of the figure, but should be typed double spaced on a separate sheet of paper. Each figure should be marked *lightly* on the back with the figure number, author's name, complete address, and shortened title of the paper.

For figures, the originals with two clearly legible reproductions (to be sent to referees) should accompany the manuscript. In the case of photographs, three glossy prints are required, preferably 8 × 10 inches.

Organization: In general, papers should state the background and purpose of the study, followed by details of methods, materials, procedures, and equipment. Findings, discussion, and conclusions should appear in that order. Appendixes may be employed where appropriate for extensive lists, statistics, and other supporting data.

Bibliography: Accuracy and adequacy of the references are the responsibility of the author. Therefore, literature cited should be checked carefully with the original publications. References to personal letters, abstracts of verbal reports, and other unedited material may be included. If an as-yet-unpublished paper would be helpful in the evaluation of a manuscript, it is advisable to make a copy of it available to the Editor. When a manuscript is one of a series of papers, the preceding member of the series should be included in literature cited.

Citation Format:

Order: Literature cited should be sequentially numbered as cited.

Authors: Give all authors with arrangement as follows:
Elias, A. W., B. H. Weil, and I. D. Welt

Titles: Give full titles of articles in English, indicating language of original as: (In Ger.)

Journals: Journal titles should be given in full.

Monograph and Serial Data: Should be presented in order as follows: Volume, issue number, pagination, and year. The issue number should be given in parentheses if journal pagination is not continuous from issue to issue. Pagination should be inclusive. Year of publication should be given in parentheses. An example is given below:

Bishop, D., A. L. Milner, and F. W. Roper, Publication Patterns of Scientific Serials, American Documentation, 16 (No. 2): 113–21 (1965).

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), three copies in the contributing membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts, Science Citation Index, Chemical Abstracts,* and *Documentation Abstracts.*

American Documentation is entered for second class mailing at Baltimore, Maryland.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 17, No. 3

JULY 1966

Published quarterly by AMERICAN DOCUMENTATION INSTITUTE and printed by Port City Press, Inc., 1323 Greenwood Road, Baltimore, Maryland 21208. Second class postage paid at Baltimore, Maryland.

Documentation Abstracts

. . . . . is a joint publication under the auspices of the American Documentation Institute and the Chemical Literature Division of the American Chemical Society.

. . . . . represents combined coverage of the former Literature Notes section of American Documentation, the ACS Division of Chemical Literature Annotated Bibliography, and the former coverage of Documentation Digest.

. . . . . will issue quarterly—February, May, August, and November of 1966. Each issue will contain corporate and author indexes; subject indexes will be available on a schedule to be determined.

Subscriptions are sold on a calendar year basis—$8.00 per year.* Return the coupon below. Payment with your order is requested.

* Members of the American Documentation Institute will receive the first year's subscription free.

DOCUMENTATION ABSTRACTS—Please enter my subscription for one year commencing with the February 1966 issue. At $8.00 per year, payment is
enclosed □ bill me □

Name____________________ Title________________

□ Business
Address □ Home ____________________________

City ________________ State__________ Zip______

Your Firm Name ____________________________

Editorial

Twenty-one years after publication of "As We May Think," considered the seminal publication for the field of Documentation, it is interesting to review the statements and positions it makes. In a recent conversation Dr. Bush noted with regret that the "MEMEX" idea remains far short of accomplishment. Knowing, as we do, that many of the devices and techniques suggested in "As We May Think" (copying techniques, instant photography, microforms, associative memories, character recognition, etc.) are in being, why does he feel this way?

The answer may well lie in the spirit of our inquiries and our research. Dr. Bush suggests more, much more than gadgetry in "As We May Think." The concluding paragraphs of that paper which follow are well worth the moments they take to read:[1]

> In the outside world, all forms of intelligence, whether of sound or sight, have been reduced to the form of varying currents in an electrical circuit in order that they may be transmitted. Inside the human frame exactly the same sort of process occurs. Must we always transform to mechanical movements in order to proceed from one electrical phenomenon to another? It is a suggestive thought, but it hardly warrants prediction without losing touch with reality and immediateness.
>
> Presumably man's spirit should be elevated if he can better review his shady past and analyze more completely and objectively his present problems. He has built a civilization so complex that he needs to mechanize his records more fully if he is to push his experiment to its logical conclusion and not merely become bogged down part way there by overtaxing his limited memory. His excursions may be more enjoyable if he can reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand with some assurance that he can find them again if they are important.
>
> The applications of science have built man a well-supplied house, and are teaching him to live healthily therein. They have enabled him to throw masses of people against one another with cruel weapons. They may yet allow him truly to encompass the great record and to grow in the wisdom of race experience. He may perish in conflict before he learns to wield that record for his true good. Yet in the application of science to the needs and desires of man, it would seem to be a singularly unfortunate stage at which to terminate the process, or to lose hope as to the outcome.

ARTHUR W. ELIAS
Editor, *American Documentation*

[1] V. Bush, As We May Think, *The Atlantic Monthly:* 101–108 (July 1945).

Coming of Age in Academe—Information Science at 21

JOSEPH C. DONOHUE and N. E. KARIOTH

Informatics, Inc.
Sherman Oaks, California

The little tyke was conceived during the fleeting affair between that somewhat shabby and passive old maid, librarianship, and the rich playboy, science. The poor kid hasn't even been named yet; in fact most of us don't even know what he looks like. Still, they were all talking about him at the FID (International Federation for Documentation) Congress in October in Washington. Some of the family felt that he ought to be reared by his mother in her flat on the outskirts of Academe. The others said that his father could give him a better life, with prestige, wealth, and status, on the best street in town. They were already calling him by his father's name, hoping he would be known as "Information Science." (1)

Information science was fathered, not by "the rich playboy, science," but rather by the concern for information transfer. It was, to be sure, born in the house of science. Its birth trauma occurred during World War II, and the birth announcement appeared in *Atlantic* in July 1945 (2). "As We May Think" has, in 21 years, generated so much thought and action that its author, Vannevar Bush, might himself be called the father of information science.

Bush, more than anyone else in this country, had been concerned with ways to facilitate the transfer of scientific information toward the very practical goal of winning the war. This done, he turned to the extension of these means to the service of research more broadly, He called for the creation of new tools, using then-existing technology, which would free the researcher from enslavement to repetitive rote operations. The system he envisioned, the "Memex", would be devoted to expanding the memory and the other intellectual processes of the researcher. Little has been done in these 21 years toward making the Memex a reality. But the concept Bush proposed, of a technology to serve the intellectual processes, gained momentum. What began as technology has become science *and* technology. Berry is right in calling information science a "tyke" only in the sense that the name is new.

A few annual meetings ago (1963), the "documentalists" of the American Documentation Institute gathered under the banner of "Automation and Scientific Communication." The next year the theme was "Parameters of Information Science." The change in emphasis from science information to a science *of* information has left some observers wondering where ADI stands. There are those, including some members, who believe that no true science will emerge, that the field will always remain a collection of disparate disciplines coming together to solve practical problems. However, this year's annual meeting theme sounds a far more optimistic note—"Progress in Information Science and Technology."

No one would contend that information science is a mature discipline. Berry reminds us [with a nod to Hoselitz (3)] of three conditions attending the birth of a new discipline:

> *Problems:* The existence and recognition of a set of new problems that attract the attention of several investigators.
> *Generalizations:* The collection of sufficient data to allow promulgation of generalizations with broad enough scope to focus on the common features of the problems under investigation.
> *Recognition:* The attainment of official or institutional recognition of new disciplines.

The *problems*, while still ill defined, are certainly recognized.

The *generalizations* do not come easily, and probably should not. But a great deal of data exists, and the data base is growing. The race to find solutions to pressing practical problems has delayed the development of unifying principles, but there is increased interest in such unification. This is evidenced by the creation of the *Annual Review of Information Science and Technology*, the first edition of which will be available at the ADI October annual meeting. This ADI project, supported by the National Science Foundation and System Development Corporation, has brought together for critical review the literature of 12 areas of special interest:

Professional Aspects of Information Science and Technology
Information Needs and Uses
Content Analysis, Specification and Control for Document Retrieval Systems
File Organization and Search Techniques
Automated Language Processing

Evaluation of Indexing Systems
New Hardware Developments
Man-Machine Communication
Information System Applications
Library Automation
Information Centers and Services
National Information Issues and Trends

As a further effort toward deriving general principles, the annual meeting will feature progress review sessions, based on these subject fields. The sessions will be panel discussions, with the chapter authors as principal reviewers.

The third condition, official *recognition*, has been extended to the problem, but not to the discipline. Vannevar Bush's position during the war is evidence of this. From the 1958 Baker Report (4) to last year's COSATI study (5), blue-ribbon panels have recognized the need for a cadre of specialists; and some have in effect advocated the creation of the discipline, but none has treated either as an accomplished fact.

Institutional recognition is shown in the growth in the number of universities now giving courses and degrees in information science—a growth so rapid that a recent editorial described it as "the education explosion" (6).

At the 1964 ADI Annual Meeting a large number of contributions were concerned with education (7), and one paper addressed itself to the problems of accreditation for the schools teaching information science.

The custodial disagreement over "information science" seen by Berry is a sign that several segments of the intellectual community find the "new science" a desirable member of their families. Fortunately, ideas need not be the exclusive property of any one group; and while the fight over custody will, no doubt, continue, information science will probably split off into various disciplines. Some segments have found homes in schools of science and technology like Georgia Tech; some, in schools of librarianship like Western Reserve and UCLA. It seems appropriate that the theoretical aspects appear to be prospering in a philosophy department at Lehigh. The philosophers have, after all, shown a strong interest in the logical and epistemological aspects of this "new" field.

Many of the schools have established ADI student chapters, which will participate in this year's national meeting (and those of the future), with a special program to be inaugurated for student members, including the award of student prize papers; these will parallel the awards made for outstanding papers contributed by the regular members.

To Hoselitz's criteria, as reported by Berry, we might add three more, as measures of maturity in a discipline.

One is responsible judgment. In the post-World War II days, emphasis shifted away from the small system that Bush envisioned to serve the individual researcher. The accent of information technology was placed on the creation of large, centralized files, such as those of the defense and space agencies.

There was what we now consider an almost touching naïvete about the ability of the giant computer to solve the problems of information transfer. A few voices of dissent were heard even during the heyday of centralization. Douglas Engelbart (8) in 1960 expounded the idea of microdocumentation, urging further development of small systems to aid the individual, much along the lines of the Bush concept.

Having pioneered in development of large systems, the Defense Department some time ago began seriously to question their value; and the current DOD emphasis on decentralization and specialization is the result. The computer-based large file at Defense Documentation Center now does what it can do well, which is secondary distribution, while information centers hold out promise of performing well a wide range of information services.

Those services are related more directly to the complex intellectual operations that are part of research. In this, they are more in keeping with the spirit of Bush's initial plan—technology in the service of the individual researcher—and they bring to the researcher the very real benefits of large-file operations.

The return to the needs of the working researcher is evident in a resurgence of interest in use studies. These are nothing new, of course. Librarians have been making them for years, usually in connection with attempts to determine the superiority of one form of catalog over another. Mortimer Taube (9) had no use for such studies, which to him made as much sense as the doctor asking the patient what treatment the patient wanted; but the current interest in use studies seems little affected by Taube's arguments. Concern with the empirical study of information use patterns, rather than with easy generalization and a priori system-building, is a sign of health in a discipline. It is probably a sign of youth, but hardly of infantilism. It must be accepted as a sign of considerable maturity that information scientists and technicians now show a tendency to abandon the belief in the easy answer, and again address themselves to the complexity of information transfer.

A second added criterion of a discipline's maturity is the adequacy of its educational structure. For information science, drawing on many subjects, it is difficult to know what should be taught, and by whom. Graduate schools are now experimenting with many kinds of curricula, but course-structuring is difficult for specialists whose backgrounds differ widely. Several ADI chapters are offering short courses. The 1966 Annual Meeting will itself present such an effort, in the form of tutorial sessions on six general subject areas:

Information System Design
Information Center Operations
Usage of Information
Evaluation of Hardware and Software
Language Data Processing
Development of a Theory

A third mark of a scientific discipline is the free flow of ideas under critical scrutiny. Research in information systems has in the past had a certain aura of the occult, for many reasons: perhaps the most important two are security regulations and the poor definitions of information system concepts. The obscurantism that resulted has sometimes been deliberate, sometimes an honest failure in communication; in either case it has stultified criticism. The promotion of a more critical attitude will be enhanced by the publication of the *Annual Review* and the presentation of the progress review sessions. Increasing educational opportunities have already had this effect.

Until recently, one potential but reluctant source for valuable criticism was the librarians; for the most part, they hung back. Some held in contempt what they considered a barbaric new technology. Some feared it, and some still do. But more and more of them have come to understand the technology, or at least what they need to understand in order to use it. As information science deepens its theoretical basis, the gulf between it and library science grows narrower, particularly with the special and technical librarians in information technology, who design and run many of the systems. At the ADI meeting in October, the Special Libraries Association will sponsor a session on "User Reactions to Non-Conventional Systems" designed to promote understanding and encourage criticism.

Such results are encouraged by the fact that, within the past year, the American Library Association has created a Division of Library Automation and Information Science. With three-quarters of ALA-accredited library schools now teaching courses in information science, the librarians in general, rather than only technical librarians, are finally beginning to engage in a dialog that will certainly improve both the free flow of ideas and the critical scrutiny of those ideas.

Information Science has gone beyond its childlike preoccupation with electronic toys. The problems it has identified are generally recognized; even the popular journals now discuss the "information explosion." The information scientists have amassed a considerable body of knowledge about the properties, behavior, and flow of information. Official and academic recognition of the discipline is in the offing.

Twenty-one years ago Vannevar Bush announced to the world the birth of a new science. The American Documentation Institute, at its Twenty-Ninth Annual Meeting to be held in Santa Monica, California, October 3–7, 1966 will point with pride to the "Progress in Information Science and Technology" that has taken place since this historic announcement.

References

1. Berry, J., It's a Wise Child, *Library Journal:* 4724 (Nov. 1, 1965).
2. Bush, V., As We May Think, *Atlantic:* 101–108 (July 1945).
3. Hoselitz, B., *Reader's Guide to the Social Sciences,* The Free Press, Glencoe, Ill., passim, 1961.
4. President's Science Advisory Committee, *Improving the Availability of Scientific and Technical Information in the United States,* U. S. Government Printing Office, Washington, D. C. 1958.
5. Federal Council for Science and Technology, Committee on Scientific and Technical Information (COSATI), *Recommendations for a National Document Handling System in Science and Technology,* 1065 (PB 168 267), The Clearinghouse, Arlington, Va., 1965.
6. Elias, A. W., Editorial, *American Documentation,* **15** (No. 2): 81 (1964).
7. American Documentation Institute, *Parameters of Information Science,* The Institute, Washington, D. C., pp. 31–77, 1964.
8. Engelbart, D. C., Special Considerations of the Individual as a User, Generator, and Retriever of Information, paper presented at the Annual Meeting of the American Documentation Institute, Berkeley, California, October 23–27, 1960, *American Documentation,* **12** (No. 2): 121–124 (1961).
9. Taube, M., *An Evaluation of Use Studies of Scientific Information,* Documentation, Inc., Bethesda, Maryland, AFOSR TN 58-1050, AD 206 987, Dec. 1958.

Librarianship and the Science of Information[*]

Utilizing query responses from accredited library schools, courses offered have been analyzed to determine emphasis on various subjects, with special reference to Information Science (IS)—i.e., the study of the information processes—and Science Information (SI)—the bibliography of science, operation of scientific information services, etc. Generally, schools continue to stress subjects related to operation of public and school libraries. Certain library schools are developing programs and/or courses in IS and SI, and in some cases the two are combined. Many schools now offer some training in nontraditional techniques. Increased stress on science information and the addition of courses in the new information technology have not radically altered the theoretical structure of library education. But the Information Science approach promises to contribute greatly to that structure. Librarianship possesses a long-established corpus of knowledge relating to a science of information. Librarians, traditionally service-oriented rather than research-oriented, have not exploited that body of knowledge for general principles of IS. Research should be directed to the areas of:

1. Cataloging and classification—the logical and epistemological underpinnings of respective systems; the relationships of given systems to prevailing theory of knowledge.
2. The technique of reference service—for its rich potential contribution to the theory of problem solving.
3. The understanding of the social and institutional framework of the information community. Library schools need now to accompany technical instruction with research into these and similar problems in order to contribute to the theory of IS and to gain from IS better technique for the practice of librarianship.

JOSEPH C. DONOHUE[†]

Informatics, Inc.
Sherman Oaks, California

This is the report of a continuing study of information science education in graduate library schools in the U.S. and Canada. The purpose of the study is to find out what is the extent and nature of each school's involvement in information science education, and to identify, if possible, a core of theory common to the curricula, as well as differences in the approaches of the respective schools.

• Information Sciences

Information science is frequently confused with science information. For the purpose of the study science information courses (Fig. 1) are those that (like science librarianship) are concerned with the techniques of information handling as applied to science. Information science, on the other hand is, to use Taylor's definition (1):

> The study of the properties, behavior, and flow of information. It includes
> (1) environmental aspects of information and communication,
> (2) information and language analysis,
> (3) the organization of information,
> (4) man-system relationships.

• Study Approach

The approach of the study is this: using catalogs, course descriptions, and in some cases additional infor-

[*] Presented October 10–15, 1965, at the International Federation for Documentation (FID), Washington, D. C. This study was begun in November 1963 and was first discussed in a short paper in the 1964 Proceedings of the American Documentation Institute, *Parameters of Information Science.* Publication of detailed data on this subject will be delayed until the analysis of each school's curriculum can be reviewed by each school's faculty.

[†] The author wishes to express thanks to the faculties of the various schools queried for their cooperation and encouragement; to colleagues at the General Electric Company-TEMPO and The RAND Corporation, for providing technical and clerical assistance. Mr. William Way of The RAND Library contributed greatly, especially in analysis of statistical data.

SCIENCE INFORMATION COURSES

TECHNIQUES OF INFORMATION HANDLING AS APPLIED TO SCIENCE

INFORMATION SCIENCE (After Taylor)

PROPERTIES – BEHAVIOR – FLOW OF INFORMATION

INCLUDES:

A. Environmental Aspects of Information and Communication
B. Information and Language Analysis
C. Organization of Information
D. Man-System Relationships

FIG. 1.

mation supplied by deans of schools, courses were divided into 21 classes by subject matter in four major categories, as follows:

Category 1: Librarianship per se—its special techniques, and its professional and ethical aspects.

Category 2: Librarianship as applied to special types of libraries, subjects, and/or clientele.

Category 3: The information sciences—the study of information, its properties, behavior and flow.

Category 4: Courses from other disciplines—given in the school of librarianship but only indirectly related to librarianship.

Only gross figures and some highlights will be presented in this report. The tabulation in Fig. 2 shows that of 34 schools finally included, course offerings total about 2950, with a mean of 87 hours, a range from 50 to 170.

In Categories 1 and 2 shown in Fig. 3, librarianship per se and librarianship as applied, the total, 93%, is about equally divided between them. The preponderance of courses in these categories is not surprising.

In Fig. 4, Category 3, the information sciences, account for 5%, and Category 4, other disciplines, totals 2%.

Schools giving at least one course in information science represent 77% of all schools. This is a significant growth from the previous year, of 30%, which is shown in Fig. 5.

With regard to the number of credit hours in courses in information science given in the individual schools, two modes are found. Eight schools offer a total of 11 hours, or about three courses; an additional 12 schools offer three hours, or one course. Most schools are approaching information science very gradually, with eight schools accounting for 75% of all courses. Examining

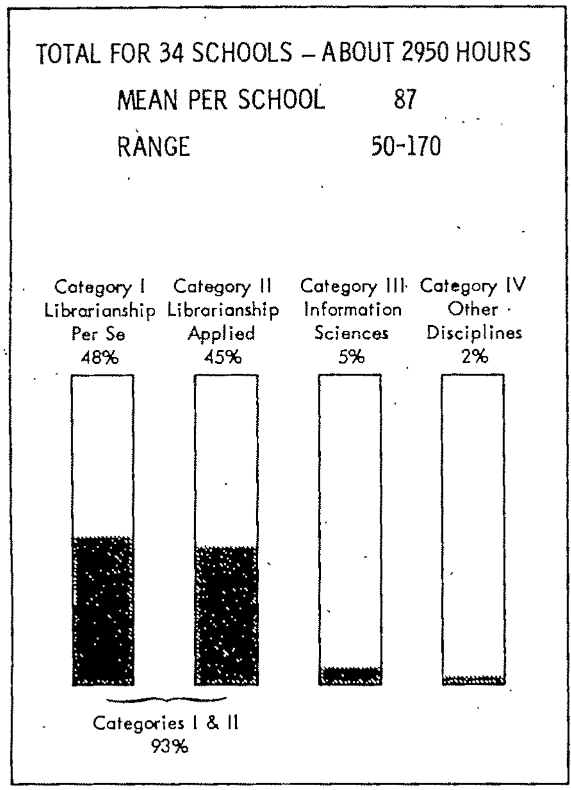


FIG. 2.

CATEGORY I LIBRARIANSHIP PER SE		Percent of All Courses
1.1	Background, History, Etc.	14
1.2	Administration	3
1.3	Selection and Acquisition	4
1.4	Cataloging and Classification	10
1.5	Technical Processes	4
1.6	Reference, Bibliography	13
	TOTAL	48%
CATEGORY II LIBRARIANSHIP AS APPLIED		
2.1	School, Children's Libraries	18
2.2	Public	6
2.3	Academic, Research	4
2.4	"Special" Libraries	2
2.5	Fine Arts	2
2.6	Humanities	4
2.7	Social Sciences	4
2.8	Science & Mathematics	5
	TOTAL	45%

FIG. 3.

CATEGORY III INFORMATION SCIENCES		Percent of All Courses
3.1	Use and User Studies	1
3.2	Operation of Information Systems	3
3.3	Theoretical Information Science – Analysis & Design	1
	TOTAL	5%
CATEGORY IV OTHER DISCIPLINES		
4.1	Languages	0.2
4.2	Mathematics	0.2
4.3	Science & Engineering	0.1
4.4	Social Sciences	1.5
	TOTAL	2.0%
		100%

FIG. 4.

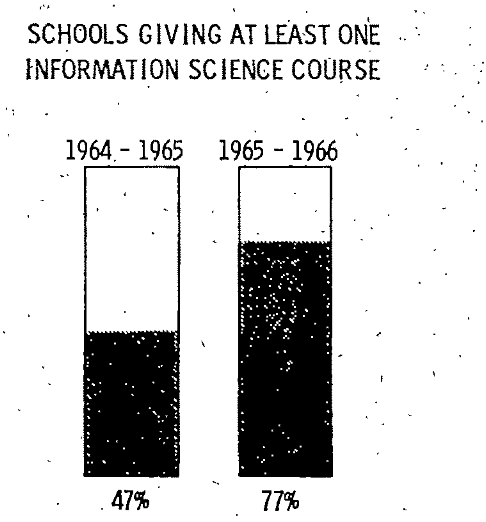
the aggregate figures for all schools, some rather dramatic contrasts are found.

In Category 1, classes in background and history lead with 14%. Reference and bibliography follow with 13%, and cataloging and classification, 10%. This distribution is especially interesting since cataloging and classification have been generally considered the major intellectual discipline of the curriculum.

In Category 2, those classes in which the basic skills are applied to particular kinds of librarianship, 18% of the total curriculum is devoted to preparation for work in children's and adolescents' libraries. This is followed by public libraries with 6%, by science—technical and mathematics (5%). Research libraries, social science libraries, and humanities libraries are about equal (4% each). Special librarianship and fine arts are in the rear guard, accounting for about 2% each.

In Category 3, the information sciences—the properties, behavior, and flow of information—the greatest number of hours are devoted to courses emphasizing operational description of nonconventional data processing and information retrieval. These courses, being for most students their only introduction to such subjects, make up 3% of the curriculum. An additional 1% is devoted to theoretical information science, including system analysis and design; and a further 1% is concerned with use studies, including study of conventional card catalog use.

Finally, in Category 4, among courses from other disciplines, given in the library school, social sciences account for 1.5%; mathematics and language, 0.2% each; and science and engineering, 0.1%.



8 Schools Offer 3 or More Courses Each or 75% of All Courses Given in All Schools

11 Schools Offer 1 Course

FIG. 5.

• **Analysis**

The relationships implicit here will require much analysis. Here are some tentative conclusions:

1. Library schools retain their overwhelming emphasis on training of personnel for public and school libraries. However, some attention is being turned to science librarianship, and to a considerably lesser extent to information science.
2. Correspondence from many of the schools indicate a strong desire on their part to strengthen the offerings in both science information and information science—but this desire is thwarted by lack of qualified instructors—in both areas, but especially in information science.
3. As demand grows and instructors become available, we may expect to see significant growth in both kinds of courses.
4. The presence of more scientifically oriented people in the library profession may, over a long period of time, indirectly affect the modes of thought of that group.
5. A more radical change may be expected as a result of the notion of a Science of Information being developed in some academic circles, including some library schools. Information science is by nature research oriented (which librarianship has not been, in this country at least). To the extent that librarianship and its schools accept that notion, it has significant poten-

tial for affecting the curriculum. One library school dean, in fact, expressed the desire that the information science approach would be reflected in every course in the curriculum, not excepting children's librarianship.

Many librarians are aware of the potential benefit of this approach. What is perhaps less appreciated is the contribution that librarianship has to offer to the theory of information science.

Recommendations

In conclusion, the following three areas require research as shown in Fig. 6:

1. *Cataloging and classification:* Implicit in the categories of classification systems, and in the techniques of cataloging are basic assumptions about knowledge and about our means of knowing. The systems should be studied as both the creators and the creations of the environments in which they arise.
2. *The structure of the information community:* Librarians, in search of information, explore and often describe in their professional literature the structure of the information community and the relations among its components. Such descriptions, properly studied, will contribute to the empirics of information science and, properly quantified can offer insights into the sociology of knowledge.
3. *Reference services:* The descriptive literature of reference service, and the skills passed on in its oral tradition contain the explanation of why a reference librarian is often more effective in literature search than our most advanced computer program. A group at Hughes Dynamics (2) has studied these techniques step-by-step showing that such analysis can offer insights on the heuristic process itself.

In these and other areas librarians can offer a great body of information, hitherto accumulated in a pragmatic and largely uncritical manner. They need now to apply analytical tools to that corpus of knowledge—to derive from it general principles, contributions to the theory of information science, and to gain in the process better tools for the practice of librarianship.

Research on THESE

1. Cataloging and Classification – Logic and Assumptions
2. Structure of the Information Community – The Sociology of Knowledge
3. Reference Services – The Heuristic Process

Will Yield

- Contributions to Information Science Theory
- Better Tools for Librarianship

FIG. 6.

References

1. TAYLOR, R. S., Review and Critique of Undergraduate Course in the Information Sciences, Report No. 1, *Curriculum for the Information Sciences,* Center for the Information Sciences, Lehigh University, Bethlehem, p. 2, 1964.
2. HUGHES DYNAMICS, INC., *Search Strategy by Reference Librarians,* Part 3 of the final report on the organization of large files, NSF Contract C-280. Sherman Oaks, California, 41 pp., 1964.

Library Catalogue Production on Small Computers*

The paper discusses the production of library catalogue cards, specifically treating the Columbia-Harvard-Yale Medical Libraries Computerization Project.† A description is provided of the bibliographic data input, general computer processing, and various output modes. The paper, also, treats of error controls, human edit procedures, and the complexity and variety of present bibliographic organization in the context of computer manipulation.

FREDERICK G. KILGOUR

Associate Librarian for Research and Development Yale University

Particulars of experiences suffered in struggles to design, write, and coax into operation computer programs to process the bibliographic data on library catalogue cards do not constitute a sizeable literature. Indeed, it is doubtful that a literature can be said to exist. Nevertheless, initial encounters with this variety of symbolic manipulative programming suggest that there is much to be learned and reported.

This paper will be constrained to a consideration of processing on general-purpose computers with final output being in upper and lower case characters. Institutions most widely known to be engaged in such activity are the National Library of Medicine (NLM) (1), Florida Atlantic University (FAU) (2), the Ontario New Universities Library Project (ONULP) (3), and the Columbia-Harvard-Yale Medical Libraries Computerization Project (CHY) (4). Stanford University and the University of California at Santa Cruz are also developing systems for book-form catalogue production.

NLM produced the first computerized issue of the *Index Medicus* in January 1964, but the printing was in the upper case characters of a high-speed drum printer. The July issue appeared in upper and lower case, having been composed on a high-speed chain printer. Since August 1964, the *Index Medicus* has been set up on NLM's GRACE (Graphic Arts Composing Equipment), which produces the most handsome computer print-out available. FAU's book-form catalogue appeared in the autumn of 1964, to be followed in the spring of 1965 by the ONULP catalogue. The first Columbia-Harvard-Yale product was a monthly accessions list that appeared in October 1963 but was in upper case characters (Fig. 6), and it was not until September 1964 that catalogue cards bearing upper and lower case characters began to be produced routinely.

Small, medium, and large computers are used to process bibliographic data. CHY employs an IBM 1401; an IBM 1460; NLM, a Minneapolis Honeywell 800; and ONULP, an IBM 7094/1401 configuration. Stanford will be using a 1401 and Santa Cruz, an IBM 1410. The Yale University Library is also designing programs to produce catalogue cards and a book-form catalogue; the catalogue-card production programs will run on an IBM 7094-7040 Direct Coupled System, but it seems likely that in the foreseeable future small computers will be much more widely available to libraries than large machines.

Moreover, the oft-repeated dictum that "the larger the computer, the cheaper the processing" does not always hold for bibliographic data processing. Some programs require prolonged reading of tapes during main-frame time; and since tapes can spin as fast, or nearly as fast, on small computers as on large, it is probable that processing is cheaper on a small machine whenever tape-spinning time equals or exceeds processing time on a large machine.

Most, but not all, of the observations reported in this paper, derive from experience with the Columbia-Harvard-Yale Medical Libraries Computerization Project. The CHY Project began in the autumn of 1961, following a suggestion by Lawrence Buckland, Ben-Ami Lipetz, and David Sparks (all at ITEK at that time)

* Work reported in this paper was supported in part by National Science Foundation Grant No. 179.

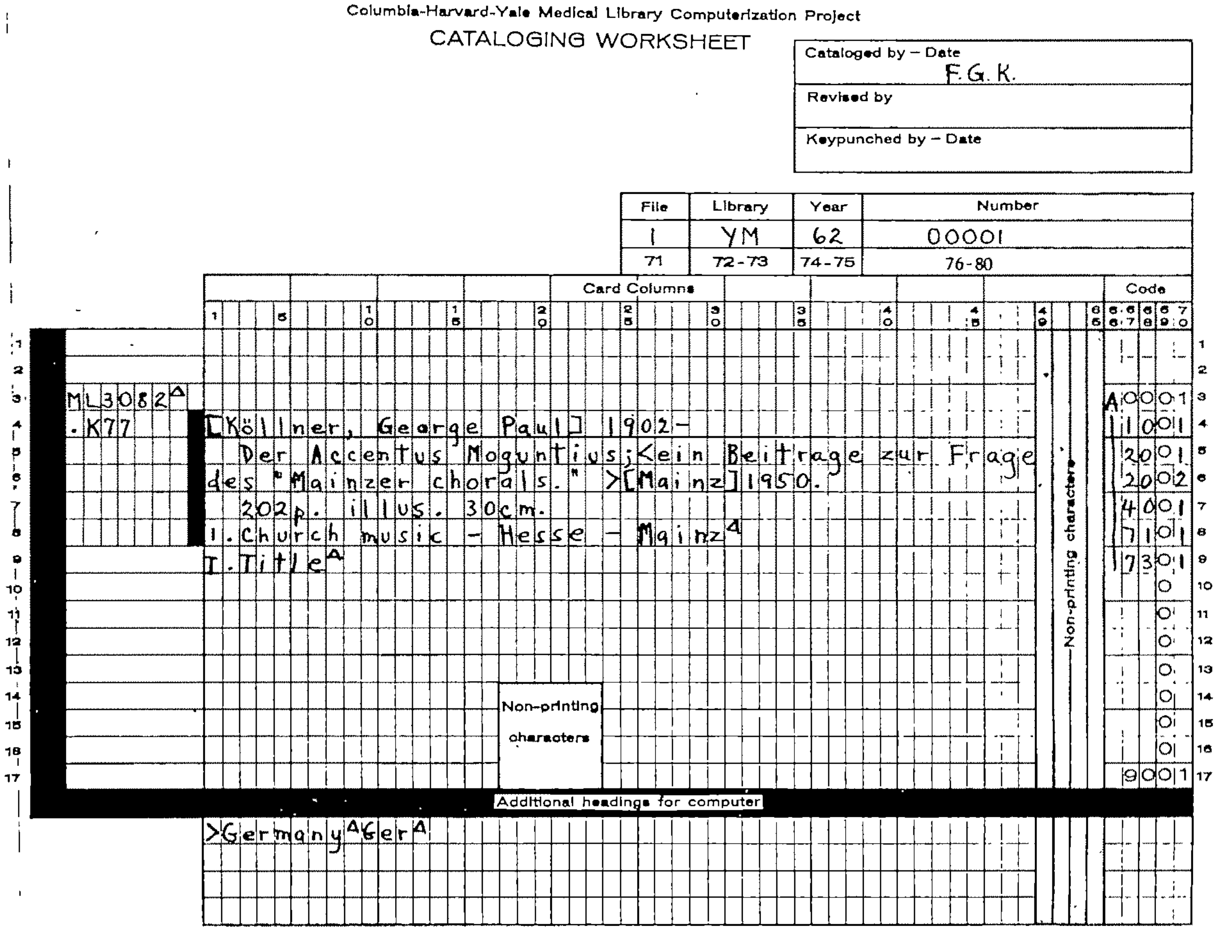
† The Harvard Medical Library withdrew from the Project as of 30 June 1966.

that it was possible to produce cataloguing information in machineable form that could be used in the production of card catalogues that could be accumulated over a period of several years to be put into a computer file when sufficient information was available to justify computerized bibliographic information retrieval. The proposal for a grant was made to the National Science Foundation in 1962, and NSF made the award in the summer of 1963. However, Yale began keypunching cataloguing information early in 1963, so that all titles processed at Yale since 1 January 1963 are in machine-readable form; the other institutions followed soon after. As already noted, CHY began routinely to produce monthly accession lists in October 1963 and catalogue cards in September 1964. Details of these procedures may be found in an article entitled "Mechanization of Cataloguing Procedures" (4).

In 1965, the Yale University Library initiated a project to computerize the procedures of the new Kline Science Library at Yale, looking forward to computerized bibliographic information retrieval. However, the system for the Kline Library is being designed to encompass all libraries at Yale. Yale is also designing an acquisitions and in-process control system that will be computerized. The aim of this second system is to gain control over all processes from the time a requester asks the Library to purchase a book until cards are in the catalogue and the book is on the shelf. The first products of these two new systems are expected to appear in the late spring and early summer of 1966, but considerable experience has already been garnered.

In the CHY system, the cataloguer catalogues each title on a worksheet (Fig. 1). The next person in the work flow (Fig. 2) is the keypuncher who punches the information on the sheet into punched cards. There is one punched card for each line on the worksheet, and the group of cards for each worksheet is called a decklet. These decklets are run through an IBM 870 Document Writer and listed off in upper and lower case for proofreading. After proofreading has been accomplished and corrections completed, groups of decklets are taken to the computer for processing; the computer is an 8K 1401 with two tape drives and modified to drive a 120-character, upper and lower case chain on the IBM 1403 printer, although the programs were originally designed for a two-tape 4K 1401.

Five programs do the processing, and each loads its successor into the computer after it has completed its



Columbia-Harvard-Yale Medical Library Computerization Project

CATALOGING WORKSHEET

Cataloged by – Date F. G. K.

Revised by

Keypunched by – Date

File	Library	Year	Number
1	YM	62	00001
71	72-73	74-75	76-80

		Card Columns	Code	
1				1
2				2
3	ML3082^		A0001	3
4	.K77	[Köllner, George Paul] 1902–	1001	4
5		Der Accentus Moguntius; ein Beitrage zur Frage	2001	5
6		des "Mainzer chorals." >[Mainz] 1950.	2002	6
7		202p. illus. 30cm.	4001	7
8		1. Church music – Hesse – Mainz^	7101	8
9		I. Title^	7301	9
10			0	10
11			0	11
12			0	12
13			0	13
14			0	14
15			0	15
16			0	16
17			9001	17

Non-printing characters

Additional headings for computer

>Germany^Ger^

Fig. 1. Cataloguing worksheet

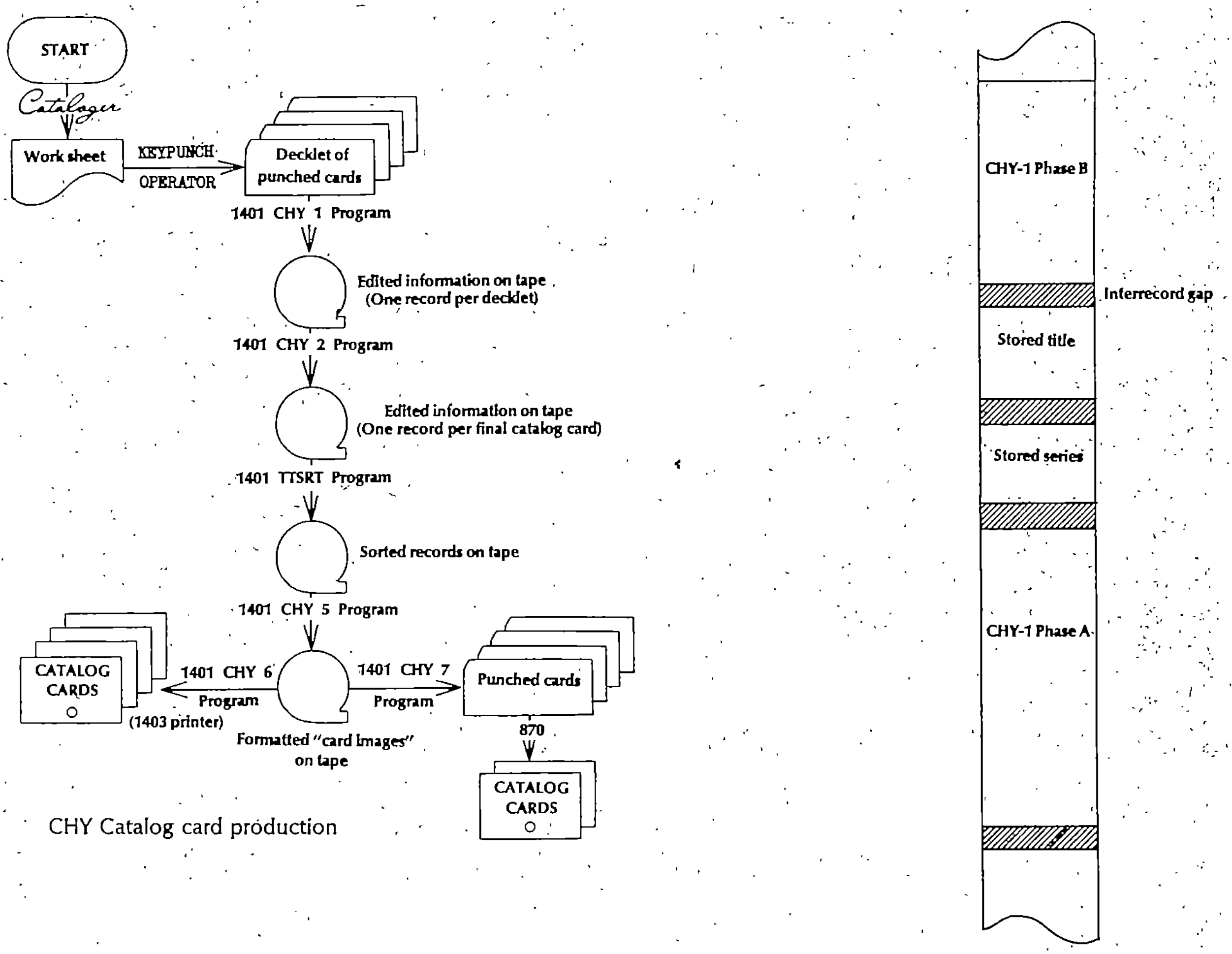


FIG. 2. Flowchart of catalogue card production

processing of the data. The first program (CHY-1) edits the data on each worksheet and writes the edited data onto a magnetic tape; it also sets up a sort control for each entry under which the card is to be filed in the card catalogue. The second program (CHY-2) explodes the edited tape record produced by the first program into a total number of tape records equivalent to the number of catalogue cards that will be needed. The third program (TTSORT) is a modified IBM package program, and it sorts the card records into various packs, each pack being destined for an individual card catalogue. The fourth program (CHY-5) sets up tape images of each card in its final format, and the fifth program (CHY-6) prints out the cards on card stock directly on the 1403 printer (Fig. 5). If catalogue cards are to be produced on an 870 Document Writer (4) instead of on a 1403 Printer, a sixth program (CHY-7) replaces CHY-6. This program punches cards on the IBM 1402 card read punch, and these cards drive the 870. When punched cards for an 870 are to be produced, the programs can be run on a two-tape 4K machine, but CHY-6 requires an 8K core as presently written.

Each month, the decklets that have accumulated during the month are run on another program to produce a monthly accessions list; on two occasions, working copies of author-entry book-form catalogues have been produced. However, like the accessions list program, the book-form catalogue program was written before the upper and lower case chain was available, so the output is all in upper case.

These programs were written in IBM's Symbolic Programming System (SPS), a symbolic-assembly language of the type to be preferred for coding programs to process bibliographic data on small computers. With these machine-oriented languages it is possible to write efficient and sophisticated programs that can take full advantage of the capabilities of a small computer or, for that matter, of a large computer. And since runs are made almost daily, it is important that processing be efficient to keep costs low.

CHY-1, the first program, edits each decklet and writes a record on tape (Fig. 3a). It also sets up the sort control by which each card can be alphabeted for filing within its pack, and writes each sort field as a separate record following the long card record. CHY-1 formulates the address of the heading in the long record at the begin-

```
a        1YM62C0001
              A00'M'L3082 .'K77‡  A10'4'K'*OLLNER, 'GEORGE 'PAUL'5 1902'/‡  A20  'DER 'ACCENTUS 'MOGUNTIUS'. EIN
             'BEITRAGE‡  A20ZUR 'FRAGE DES '1'MAINZER CHORALS.'1  '4'MAINZ'5‡  A201950.‡  A40  202P.  ILLUS.  30
            CM.‡  A711.'CHURCH MUSIC '/ 'HESSE '/ 'MAINZ ‡  A73'I.'TITLE ‡


                  0211KOELLNER GEORGE PAUL 1902‡
              2H2 0411CHURCH MUSIC - HESSE - MAINZ‡
                  1U0J1011ACCENTUS MOGUNTIUS‡


b        03 ACCENTUS MOGUNTIUS‡                                  1YM6200001     1UC  A00'M'L3082 .'K77‡  A10'4'
         K'*OLLNER, 'GEORGE 'PAUL'5 1902'/‡  A20  'DER 'ACCENTUS 'MOGUNTIUS'. EIN 'BEITRAGE‡  A20ZUR 'FRAGE D
         ES '1'MAINZER CHORALS.'1  '4'MAINZ'5‡  A201950.‡  A40  202P.  ILLUS.  30CM.‡  A711.'CHURCH MUSIC '/
         'HESSE '/ 'MAINZ ‡  A73'I.'TITLE ‡


c        =CC
         =10'DER 'ACCENTUS 'MOGUNTIUS
         =C0'M'L3082
         =C0.'K77     '4'K'*OLLNER, 'GEORGE 'PAUL'5 1902'/
         =08  'DER 'ACCENTUS 'MOGUNTIUS'. EIN 'BEITRAGE
         =08ZUR 'FRAGE DES '1'MAINZER CHORALS.'1  '4'MAINZ'5
         =081950.
         =08  202P.  ILLUS.  30CM.
         =00  'CT'Y'/'M 62    1
```

FIG. 3. (a) Tape dump of records written by edit program, CHY-1. Three sort fields are at the bottom. (b) Tape dump of record produced by CHY-2. First two digits are pack number followed by alpha sort field, identification codes, and address (1U0) of title with an A bit on the middle digit to indicate heading is an added entry. (c) Tape dump of card image set up for printing

ning of each sort field, except for the main entry. By placing bits on the middle digit of the address, it indicates whether the heading is a topical subject, name subject, or added entry, information that CHY-5 will need to determine the location of each type of heading on its respective catalogue card. However, the programs have been run on two-tape configurations on which it has not been possible to sort alphabetically. As a result, the subroutines for setting up the sort fields have not been completely debugged.

CHY-1 requires over 5,900 spaces of core and in its 4K version is stored in part on tape. Some 1,600 spaces provide a work area, and place for control and processing routines. Two groups of subroutines, Phase B and Phase A, of 2,388 and 1,953 characters are written on tape to be called into core to overlay each other as required. The program also writes the title and series category of each decklet between the two subroutines, because these categories of data must be retrieved if a sort field for a title, short title, or series entry is to be set up. Since these data are of varying lengths, the Phase A must be rewritten each time before the first is called (Fig. 4).

The most difficult programming for processing bibliographic data is the setting up of sort controls. The complex requirements for writing sort fields determined the size of CHY-1. As Fig. 3a shows, punctuation and flags are removed except for dashes; in the main entry in the illustration, the square brackets, indicating that the author's name is not on the title page, have been removed as well as commas. The program converts ampersands to the correct conjunction for the language of the title page and transliterates, if necessary, characters with diacritical marks into the equivalent letters in the filing alphabet. For instance, the "ö" in the author's name illustrated becomes "OE" in the sort field. It is also necessary to remove an initial article from the sort field as shown in the figure. These and similar manipulative routines are

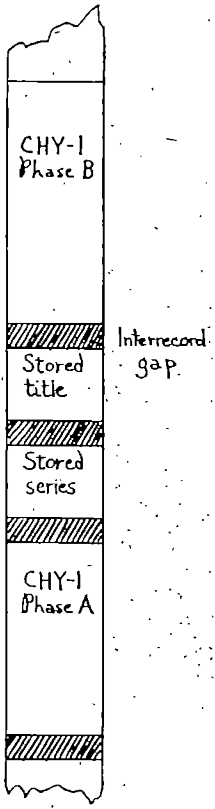


FIG. 4. Map of CHY-1 system tape

space consuming and for book-form catalogue production must be much more accurately designed than for the production of catalogue card packs.

In the original CHY sequence of programs that produce punch cards to drive an 870 (4), CHY-1 was followed by three programs—CHY-2, TTSORT, and CHY-4. However, when the upper and lower case chain became available, for which CHY-4 had not been designed, it was necessary to rewrite that program in two sections. The design for CHY-2 also proved inadequate and was rewritten. In their original form, CHY-2 and CHY-4 contained generator programs that wrote short series of a half-dozen instructions at most constituting several programs that the load program used. Data on a control card determined the character of the generated programs. This technique is efficient in use of space and may augment speed if either or both are required; a four- or six-instruction program usually yields more rapid processing than a table search and is shorter than subroutines to be initialized with addresses and constants.

The three programs that presently follow the edit program are the second edition of CHY-2, TTSORT, and CHY-5, and these three programs use about 11,500 spaces of core. Each contains an initialization or generator program varying in size from 300 to 860 characters that is subsequently overlain by a work area. As already mentioned, CHY-2 expands the records it receives from CHY-1 into the number of records equivalent to the number of 3×5 catalogue cards required for each title. It accomplishes this expansion by combining the information received from CHY-1 as to the number of heading cards required, the main-entry card, and information on the control card for CHY-2 giving the number of packs of cards and the type of cards to go in each pack. This program is capable of setting up 99 packs, each containing from one to five types of catalogue cards (main entry, topical subject headings, etc.) and will place the tracings for the headings at the bottom of each type of card as punched on the control card. CHY-2 also sets up the sort control for call numbers, assigns this control and the controls received from CHY-1 to a sort field in the new record, and assigns for each new record a pack number in which the card will ultimately be printed (Fig. 3b).

TTSORT is a slow, two-tape sort that processes digit by digit, column by column. If four tapes were available on the equipment at Yale, it would be possible to sort more rapidly by pack and to alphabetize the cards within each pack. Next, CHY-5 writes on tape the image of the card as it will be printed on the upper and lower case chain or on the 870 Document Writer (Fig. 3c).

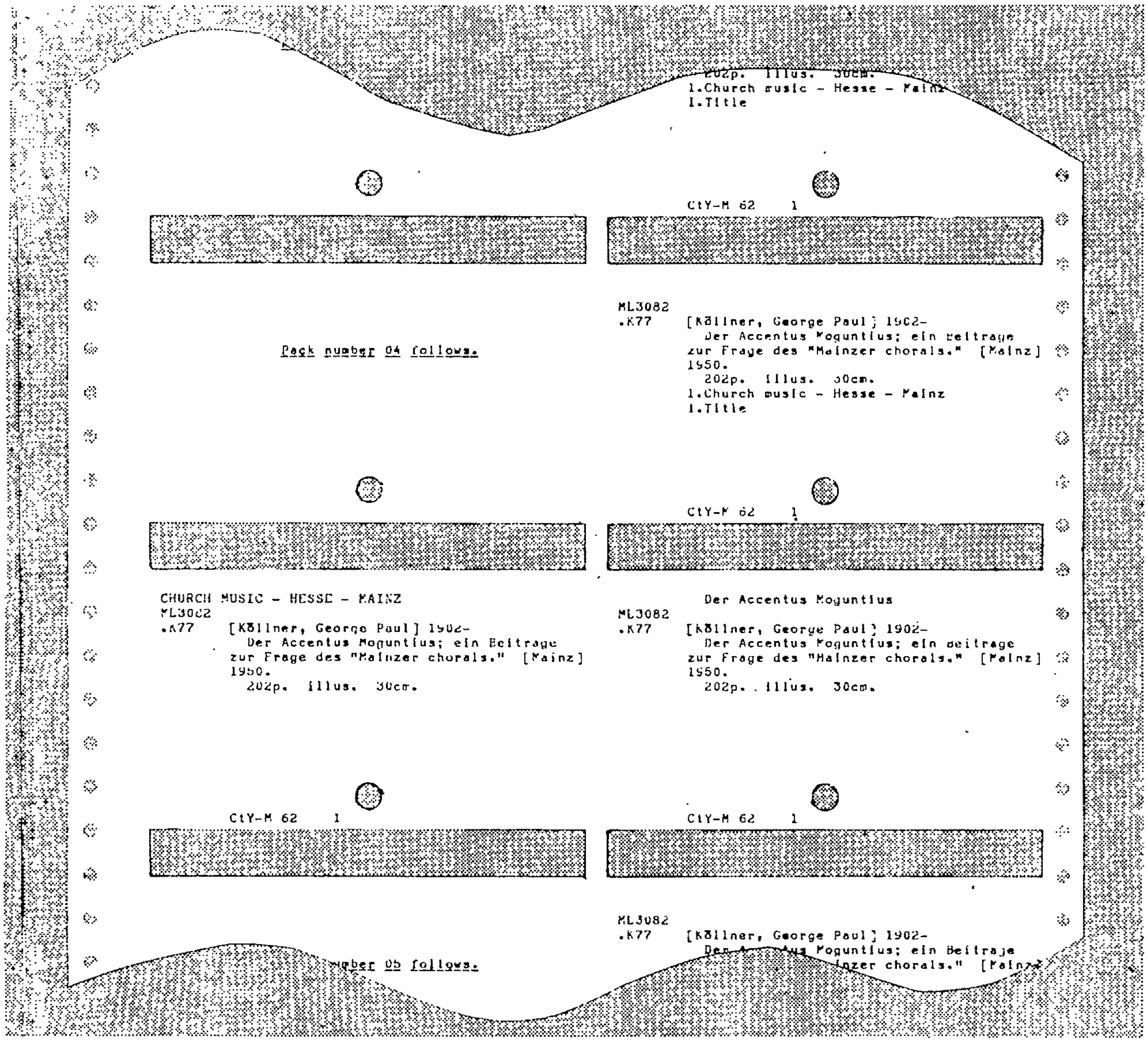
Either CHY-6 or CHY-7 follows CHY-5. As already noted, the first prints cards on the 1403 and does so either side by side as shown in Fig. 5, or one up, the choice of output being determined by the setting of a sense switch. If CHY-7 is used instead of CHY-6, cards are punched on the 1402 containing the information from CHY-5; and these cards, as already noted, drive an 870. CHY-6 uses 4,440 spaces of core, and CHY-7, nearly 1,800.

Although the programs have been operating for a year and a half, they still harbor two major bugs, each of which performs its trouble-making about once or twice during a week. Also, the programs produce only single catalogue cards and not continuation cards. Debugging is not complete because of press of other duties, while production of continuation cards will require the redesign of CHY-1.

As might be divined from this detailed description of the CHY set of programs for processing bibliographic data, the original conception of the processing was that of a series of programs, each independently loaded in sequence by the operator. In the process of writing the programs, it became clear that it was feasible to have each load its successor and reduce human intervention in processing. As presently written, the five programs used for punch-card production for 870 operation employ over 19,000 spaces of core but run on a 4K machine.

Retrospection suggests that better design might be achieved in such programs by designing them as one program which might be 19K or more, and then converting the program into modules for sequential processing on a computer with a smaller core. Of course, total design could not be done independently of modular sequences; rather, the modules would have to be roughly outlined before the processing design was undertaken. No matter what the approach, the sort-control module will be the largest component and the most difficult to design.

Experience at Yale with the programs described above and in designing further programs already mentioned for the production of catalogue cards and book-form catalogues throughout the University Library system has already yielded interesting observations. Similar observations have been made from the experience of others. Perhaps the most consistent and most striking observation is that the programming of bibliographic data without exception has taken longer, or has required more programmers, than originally estimated. Where there have been deadlines for operational production, they have not been met, although the FAU printed catalogue was very close to its deadline. At Yale there has also been experience in writing programs that produce output in upper case characters only on the 1403, as well as the type of upper and lower case output described. In programs for upper case output, the same bibliographic input data is employed as used in card production, but flags and characters not on a 48-character chain are removed or converted in the read subroutine. An example of such output is the accessions list depicted in Fig. 6. Comparison of times required to write programs for specialized output in upper case, such as the accessions list or a main entry catalogue, with time required to produce bibliographically accurate output in the form of catalogue cards or book-form catalogues yields the rather surprising result that it takes from 6 to 20 times as long to program the latter. It is this extraordinarily large differ-



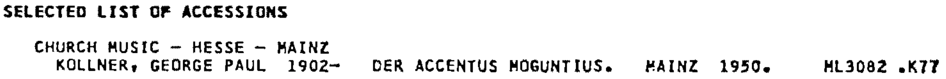
```
                                                  202p.  illus.  30cm.
                                                  1.Church music - Hesse - Mainz
                                                  1.Title



                                                  CtY-M 62      1



                                          ML3082
                                          .K77     [Köllner, George Paul] 1902-
                                                     Der Accentus Moguntius; ein Beitrage
        Pack number 04 follows.                    zur Frage des "Mainzer chorals." [Mainz]
                                                   1950.
                                                     202p.  illus.  30cm.
                                                   1.Church music - Hesse - Mainz
                                                   1.Title



                                                   CtY-M 62      1


CHURCH MUSIC - HESSE - MAINZ                          Der Accentus Moguntius
ML3082                                    ML3082
.K77     [Köllner, George Paul] 1902-     .K77     [Köllner, George Paul] 1902-
           Der Accentus Moguntius; ein Beitrage      Der Accentus Moguntius; ein Beitrage
         zur Frage des "Mainzer chorals." [Mainz]  zur Frage des "Mainzer chorals." [Mainz]
         1950.                                     1950.
           202p.  illus.  30cm.                      202p.  illus.  30cm.



         CtY-M 62      1                           CtY-M 62      1



                                          ML3082
                                          .K77     [Köllner, George Paul] 1902-
                                                     Der       Moguntius; ein Beitraje
             mber 05 follows.                                  inzer chorals." [Mainz
```

FIG. 5. Catalogue cards printed two up on an IBM 1403

ence in effort required that leads to deadlines not being met.

The question not unnaturally arises as to why it should take so very much longer to program bibliographic data processing to yield a product having bibliographic integrity. There are three major reasons: first, the bibliographic data for each title is surprisingly complex for a small amount of data; second, the determination of the location of each bibliographic entry in a large catalogue is a complex procedure; and third, the generalization of programs giving them the capability of yielding catalogue cards and book form catalogues having different formats adds another highly complex factor.

Inspection of Fig. 1 will give some appreciation of the complexity of data on a catalogue card, although at first glance the data seem quite simple. In the card shown, it was necessary to remove the punctuation (including the square brackets) to set up the sort control, and to eliminate the spaces occupied by punctuation so that the sorting field was left justified. Also, it was necessary for the program to recognize that the text being in German, the lower case "ö" with the umlaut over it, requires translit-



```
SELECTED LIST OF ACCESSIONS

   CHURCH MUSIC - HESSE - MAINZ
     KOLLNER, GEORGE PAUL  1902-   DER ACCENTUS MOGUNTIUS.  MAINZ  1950.     ML3082 .K77
```

FIG. 6. Sample output for accessions list from same decklet used to produce Figs. 3 and 5

eration into "OE" to be filed in its proper location. Similarly, the program had to recognize that "Der" was a German article to be removed from the sort control that determines the filing location of the title added entry. Moreover, the program must be able to accomplish the umlaut conversion and the article removal in author entries as well as in title, short title, and series-added entries.

From these examples and others already given, it is obvious that there is a huge, although not infinite, variety of outputs within categories of bibliographic data on each catalogue card. The variety is so enormous that it would be impossible to list each output, and moreover it would be useless to do so. In short, it is necessary to generalize the output to some extent even before flowcharting is undertaken.

The CHY programs described were designed to yield printed catalogue cards in packs and alphabeted within each pack for filing purposes, the filing, of course, to be done by humans. In the design of a book-form catalogue, it is necessary to set up the sort fields so that each entry, be it author, title, short title, conventional title, series, subject, joint author, editor, etc., will be located unambiguously in the arrangement of the catalogue so the users can find it following simple rules. Complexity of programming increases as the number of titles in the catalogue increases and as the number of types of entries for each title in the catalogue increases. For instance, in an author, book-form catalogue there may be as many as 10 different types of entries, and a single filing sequence may involve up to four factors such as surname, given names, date of birth, and a qualifying phrase describing the author.

In a system of book-form catalogues, there should be an author catalogue, a title catalogue, subject catalogue, shelf-list, and official catalogue of which the first three, or two of the first three, may be combined into one. It is this type of generalization of output that contributes the third major factor of complexity. When these three factors converge in one program, the overall complexity expands rapidly; it is this fact, more than any other, that necessitates the expenditure of a surprisingly large amount of effort to program bibliographic data processing.

Recommendations for specific details in program design appear elsewhere (5), but there are at least two more general observations to be made. Design should consciously attempt to reduce human participation to the absolute minimum necessary to avoid introduction of error. The history of machines, and particularly of machine tools, is characterized by an increasingly accurate product that has made possible assembly-line mass production employing interchangeable parts. Similarly, the more complete the machining of bibliographical records, the more universal will be the use of those records. Some programmers have relieved the computer of processing it could do and placed responsibility on human cataloguers. Thereby, programming is simplified and processing is more "positive;" but irrespective of such views, experience confirms that the machines can be trusted.

Examples of activities assigned to humans which are better handled by machine include special flags to exclude initial articles to be dropped from a sort field; leaving a space to insert a nonspacing flag; and writing dates that can be extracted from the data. Computers are entirely capable of coping with such tasks and can assume the burden. To insure that the machine assumes as much of the burden as possible, the final design of input data should receive an isolated, last review to determine that humans are doing nothing in preparation of the data that the computer can do.

Finally, there are different values of error in input bibliographic data. Some types of error can prevent processing of a decklet, and both human and machine procedures should guard against and detect such error. At the other end of the spectrum are errors that do not interrupt processing and which cannot be recognized in the final output. An example of this latter type of error might be a missing added-title entry that the cataloguer intended to put in the catalogue but for which he neglected to give the instruction. It is most likely that the prevalence of this type of error is far less than the incidence of varying judgments by cataloguers as to whether or not there should be a title added entry. There is no point in guarding against error of omission when varying judgment yields a far higher incidence of omission.

• Conclusion

Until there is a cadre of trained individuals who are equally at home in librarianship, systems analysis, and programming, it seems likely that the enormous complexities of bibliographic data processing will confine its development to multistage advancements. In fact, the advancements will probably have occurred before the cadre is in being. The Yale work is an example of multistage development, there being two major steps in the CHY Project which have formed the basis for a third major step that will be the Yale University Library's system for production of catalogue cards and book-form catalogues. In review, it seems most improbable that it would have been possible to go directly to the third step and accomplish the sophisticated output now being designed into the system; initial programs elsewhere will probably be subsequently elaborated. In the light of this observation, it seems wise at the start of a project for bibliographic data processing to recognize the probability, and perhaps even the desirability, of a multistage development.

The amount of effort required to program for bibliographic data processing is so large as to generate a pressing need for multiuse programs to be shared by different institutions. It is to be hoped that at some stage in the development of this type of processing such multi-use programs will come into being and thereby reduce programming effort. It is now obvious that although it was too much to hope for universal use of programs at the first stage, such programs should be the future objective of those working in the field.

References

1. Adams, S., MEDLARS: Performance, Problems, Possibilities, *Bulletin of the Medical Library Association,* **53** (No. 2): 139–151 (1965).
2. Perreault, J., Computerized Cataloging: The Computerized Catalog at Florida Atlantic University, *Library Resources & Technical Services,* **9** (No. 1): 20–34 (1965).
3. Bregzis, R., The Ontario New Universities Library Project—An Automated Bibliographic Data Control System, *College & Research Libraries,* **26** (No. 6): 495–508 (1965).
4. Kilgour, F. G., Mechanization of Cataloguing Procedures, *Bulletin of the Medical Library Association,* (No. 2): 152–162 (1965).
5. Kilgour, F. G., Symbol-Manipulative Programming for Bibliographic Data Processing on Small Computers, *College & Research Libraries,* **27** (No. 2): 95–98 (1966).

On the Optimum Number of Frames Per Microfiche

It is shown that, in reducing a collection of documents to microfiche, the efficiency of the microfiche-microframe system has an optimum value. This optimum depends on the page distribution of the collection. For example, the optimum for the NASA and DOD collections is 63 frames per microfiche. The ratio of pages to frames is then 0.57, and the ratio of documents to microfiche 0.72. The existence of such an optimum suggests that further significant filming reduction ratio should be accompanied by a corresponding reduction of microfiche size.

WOLF KUEBLER [1]
Documentation Incorporated
Bethesda, Maryland

• Introduction

In recent years, the use of microfiche as means of communication of information within the scientific community has been well established. The National Aeronautics and Space Administration (NASA), the Atomic Energy Commission (AEC), and the Department of Defense (DOD) are the foremost examples for the use of microfiche on a large scale.[2]

Microfiche are an economic means for the mass distribution of published literature. This economy extends to essentially three parts: economy of storage, economy of reproduction, and economy of transmission. Most recently, an effort [3] has been made to standardize the physical size of microfiche and the reduction ratio, and therefore the number of frames per microfiche. The reduction ratio is established by the state of the art of the microfilming and reproduction technique, and it can be expected that the present 18:1 ratio will in the future be superseded by a higher ratio. Thus the existence of an optimum way to store the documents of a collection on fiche is an important consideration in further standardization efforts.

• Page Distribution in a Collection of Documents

To find the page distribution, the approach is taken that each member of the community generating these documents is producing useful information (undefined), and that to express this information, a variable number of words as well as visual representations, such as pictures and graphs, are necessary. To take these latter into consideration, a page will be considered as the fundamental quantity of information, making no distinction in the subjective value of the information on each page; each page has as much information content as the next page. Furthermore, it is assumed that the page size and word density average out over a large population.

A collection of D Documents is composed of D_1 documents with 1 page, D_2 documents with 2 pages, and generally of D_p documents with p pages. The number W of possible arrangements of these D_p documents in p page classes is then

$$W(D_p) = \frac{D!}{\Pi D_p!} \qquad (1)$$

where

$$D = \sum_{p=1}^{\infty} D_p \qquad (2)$$

represents the total number of documents.

$W(D_p)$ is the discrete probability distribution for the occurrence of D_p documents with p pages. The most probable distribution is calculated, by using equation (2) and the total number of pages

$$P = \sum_{p=1}^{\infty} pD_p \qquad (3)$$

as constraints. In place of $W(D_p)$, one seeks the unconditional maximum of ln W, taking care of the accessory conditions by Lagrangian multipliers.

$$\delta\left[\ln W(D_p) - \alpha \sum_p D_p - \beta \sum_p pDp\right] = 0 \qquad (4)$$

[1] Present address: Documentation Incorporated, 4833 Rugby Avenue, Bethesda, Maryland.

[2] NASA, for example, has approximately 50,000 documents on microfiche.

[3] COSATI.

Using Stirling's formula for ln $W(D_p)$ gives

$$\sum_p \ln D_p \delta(D_p) + \alpha \sum_p \delta(D_p) + \beta \sum_p p\delta(D_p) = 0 \qquad (5)$$

This is true for every p; therefore

$$\ln D_p + \alpha + \beta p = 0 \qquad (6)$$

or

$$D_p = e^{-\alpha-\beta p} \qquad (7)$$

The Lagrangian multipliers are determined from the constraint conditions expressed by equation (2) and equation (3). One finds[4]

$$e^{-\alpha} = (e^{\beta} - 1)D \qquad (8)$$

$$\beta = -\ln(1 - D/P) \qquad (9)$$

where P/D is the average number of pages for a given collection. Finally, the page distribution is obtained

$$D_p = (e^{\beta} - 1)e^{-\beta p} \cdot D \qquad (10)$$

or

$$D_p = \frac{D}{P/D - 1}(1 - D/P)^p \qquad (11)$$

For further calculation, the result in the form of equation (10) is preferred.

Before comparing the results of this calculation with actual collections, a short discussion as to the nature of document collections is necessary. Basically, the documents fall into two classes: Class A consists of the "Open Literature"; Class B, of the "Report Literature." Class A documents are journal articles, reports of symposiums, etc., and by their nature are relatively short, say less than 30 pages. Class B documents are reports not restricted in length, although they practically are usually less than 300 pages.

The page distribution of Class A documents was obtained by a sampling of the *International Aerospace Abstracts* (*AIAA*).[5] The results are shown in Fig. 1 and in Table 1. The solid line represents equation (11), with P/D obtained from the sample. The interesting features are the low value and the overshoot on the first nine pages. This is accountable for by human elements, which were not taken into account in finding the most probable distribution. For example, these Class A documents contain the papers presented at symposiums. A certain time is "allowed" each speaker, resulting in papers clustered between 3 and 7 pages. This feature will not influence further calculations, as long as $p \geqq 10$. In Fig. 3, the result is plotted in 10-page intervals; as can be seen, the oscillating part averages out.

TABLE 1. The ratio Page to Document (P/D) for different samples.

Collection	Type	Sample size Documents D	 Pages P	P/D	β
AIAA	Open Literature (Class A)	920	7011	7.7	0 14
NASA	Open Literature (Class A)	763	5611	8.4	0.12
	Report Literature (Class B)	1448	72416	50	0.02
	82 Frame * Microfiche (Class B)	950	———	———	0.02
	58 Frame * Microfiche (Class B)	1192	———	———	0.02
DOD	Report Literature * (Class B)	1271	———	———	0.02

* The pages were not counted; the P/D value found for the NASA Class B collection was sufficient to represent the data.

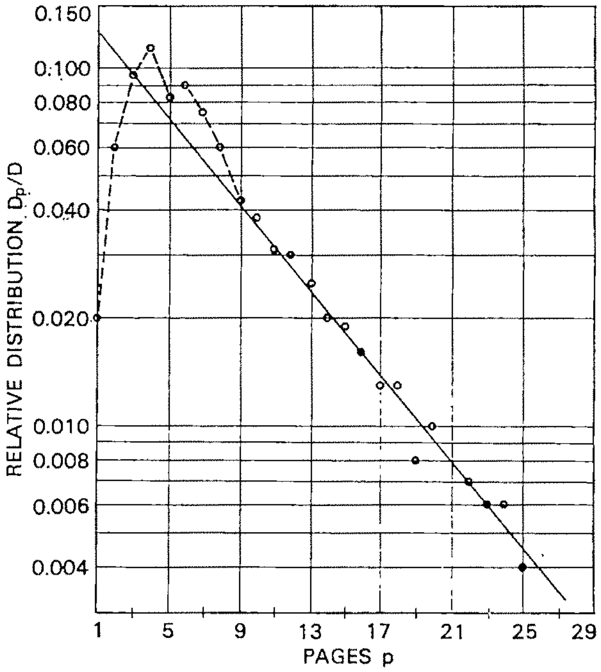


FIG. 1. Page distribution obtained by sampling *International Aerospace Abstracts* (Class A documents, open literature); the solid line represents equation (10)

To calculate the distribution function for intervals, let the document collection be subdivided into intervals of documents having pages (1, 2, . . . , q) for the first interval, (q+1, q+2, . . . , 2q) for the second interval, and ((n−1) q+1, (n−1) q+2, . . . , nq) for the nth interval. The number of documents, $D_{(n-1)q+1,\ nq}$ in the nth interval is then by equation 10:

$$D_{(n-1)q,\ nq} \equiv D_n = D(e^{\beta} - 1) \sum_{(n-1)q+1}^{nq} e^{-\beta q}$$

$$= D(e^{\beta(q-1)} - 1)e^{-n\beta q} \qquad (12)$$

This again is an exponential function, the primary variable being the interval number n.

The page distribution of Class B documents was obtained by a sampling of *Scientific and Technical Aerospace Reports* (*STAR*)[6] and *Technical Abstract Bulletins* (*TAB*).[7] The results are shown on Fig. 2 and Table 1 for

[4] The summations are from $p = 1$ to $p = \infty$. The error introduced is negligible; in principle the calculations could apply to a finite scheme.

[5] Published by the Technical Information Service, American Institute of Aeronautics and Astronautics.

[6] Published by the National Aeronautics and Space Administration.

[7] Published by the Defense Documentation Center.

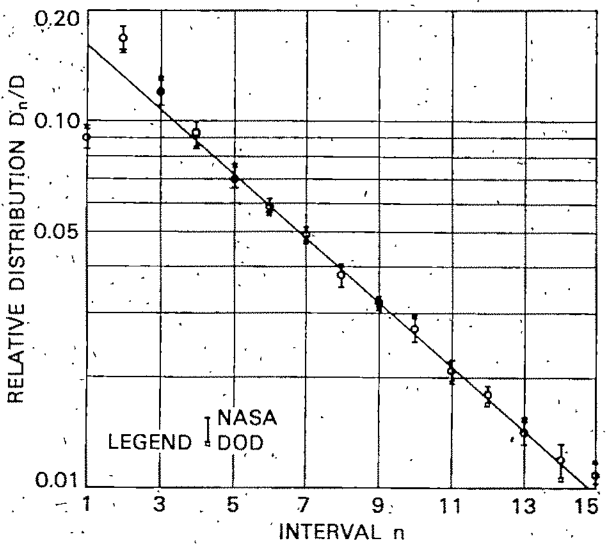


FIG. 2. Page distribution obtained by sampling *Scientific and Technical Aerospace Abstracts* and *Technical Abstract Bulletin* (Class B documents, report literature); page interval q = 10

10-page intervals.[8] The solid line represents equation (12) for q=10 page intervals with P/D obtained from the *STAR* sample. There is again an overshoot with a peak at 20 pages. In Fig. 3, the results have been plotted with 30-page intervals; the overshoot has been smoothed out, as has the "spread" between the two collections.

Finally in Fig. 3 and Table 1 are the results of a sampling of the same Class B documents: NASA microfiche. There are two sizes; previous to standardization, the microfiche were on 5"×8" size with 82 frames; now the microfiche are on 4"×6" size with 58 frames.

The results of this part are necessary in order to calculate an efficient microfiche size for documents of a given collection.

• Efficiency of a Document-microfiche Collection

The interval q is now identified with the number of frames in a microfiche. In the nth interval, there are according to equation (12), $D_{(n-1)q+1,\, nq}$ documents; consequently, there are $nD_{(n-1)q+1,\, nq}$ microfiche M_n for this interval:

$$M_n = nD_{(n-1)q+1,\, nq} = D(e^{\beta(q-1)} - 1)ne^{-n\beta q} \qquad (13)$$

The total number of microfiche M is therefore

$$M = \sum_{n=1}^{\infty} M_n = D(e^{\beta(q-1)} - 1)\frac{e^{-\beta q}}{(1-e^{-\beta q})^2} \qquad (14)$$

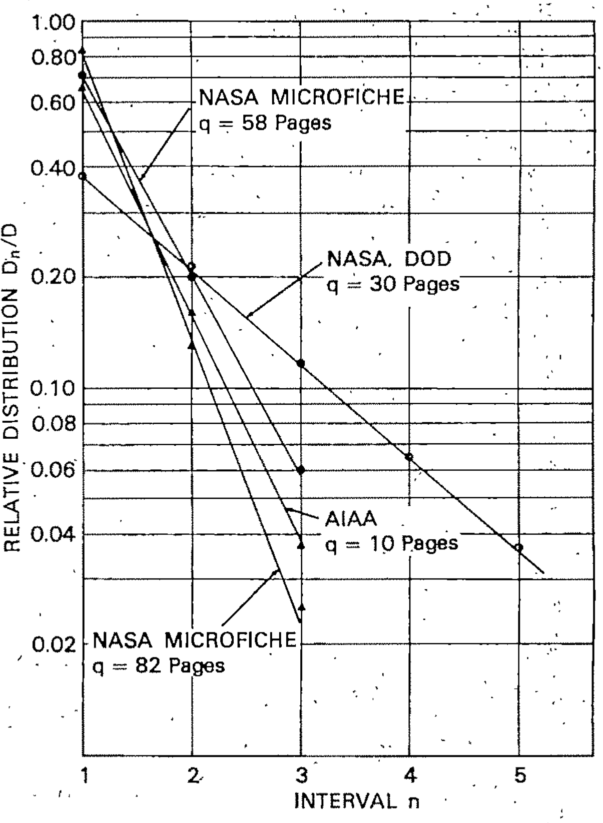


FIG. 3. Page distribution obtained by sampling microfiche (q = 58 and q = 82 pages); included are the data of Figs. 1 and 2 for different page intervals

Fig. 4 shows the relative distribution of microfiche

$$\frac{M_n}{M} = (1-e^{-\beta q})^2 e^{\beta q} n e^{-n\beta q} \qquad (15)$$

for different values of βq. Equation (15) has a maximum for $n\beta q=1$. As long as a given document collection fulfills the inequality

$$\beta q \geqq 1 \qquad (16)$$

one is assured that most documents are on one microfiche. For $\beta q=1$ approximately 63% of the documents are on one microfiche, representing approximately 40% of the total number of microfiche.

The efficiency of the document-microfiche system consists of two aspects. First there is the efficiency η_M in terms of the ratio of total number of microfiche to the total number of documents[9]

$$\eta_M = \frac{D}{M} = 1 - e^{-\beta q} \qquad (17)$$

In terms of the number of microfiche, the most efficient system ($\eta_M=1$) is obtained if $\beta q \to \infty$. Secondly, the

[8] This was a selective sampling. Whereas in the sampling of *AIAA*, each item was used, for *STAR* and *TAB* only the "open literature" items were considered. For *STAR* also the Class A items were considered; the result is listed in Table I. The sampling is represented in 10-page intervals; otherwise, to get a distribution, a sample about 10 times the size would have been necessary.

[9] For simplicity, it is assumed that q >1; i.e., the microfiche has say more than q = 10 pages. Thus equation (14) reduces to

$$M = \frac{D}{1-e^{-\beta q}}$$

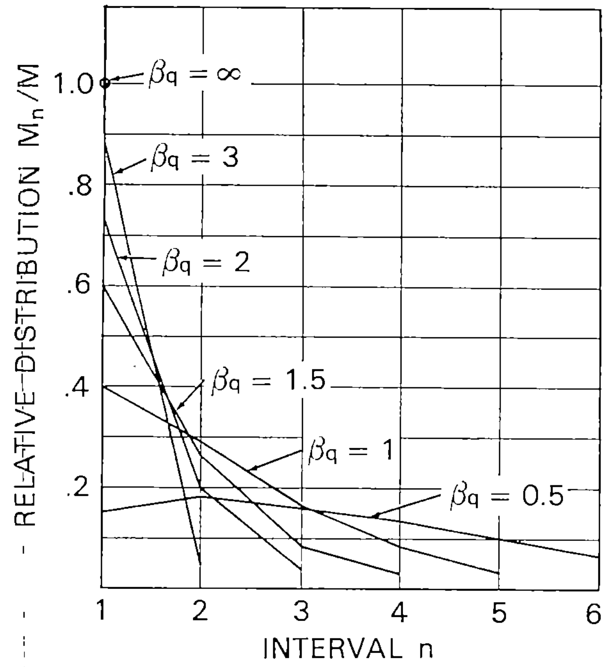


FIG. 4. Relative distribution of microfiche for different βq

efficiency η_F of the ratio of total number of pages P to total number of frames qM is given by

$$\eta_F = \frac{P}{F} = \frac{P}{qM} = \frac{1}{D/P}\,\frac{1 - e^{-\beta q}}{q} \tag{18}$$

If D/P $\lessdot$ 1, then by equation (9)

$$D/P \simeq \beta \tag{19}$$

and

$$\eta_F = \frac{1 - e^{-\beta q}}{\beta q} \tag{20}$$

In terms of the number of frames, the most efficient system ($\eta_F = 1$) is obtained for $\beta q \rightarrow 0$. The overall efficiency is

$$\begin{aligned} \eta &= \lambda \eta_M \cdot \eta_F \\ &= \frac{(1 - e^{-\beta q})^2}{\beta q} \end{aligned} \tag{21}$$

The factor $\lambda = 2.46$ is introduced to normalize the overall efficiency, the most efficient microfiche-frame system has the efficiency one. Equation (21) is shown on Fig. 5. It has a maximum for

$$(1 + 2\beta q)e^{-\beta q} = 1 \tag{22}$$

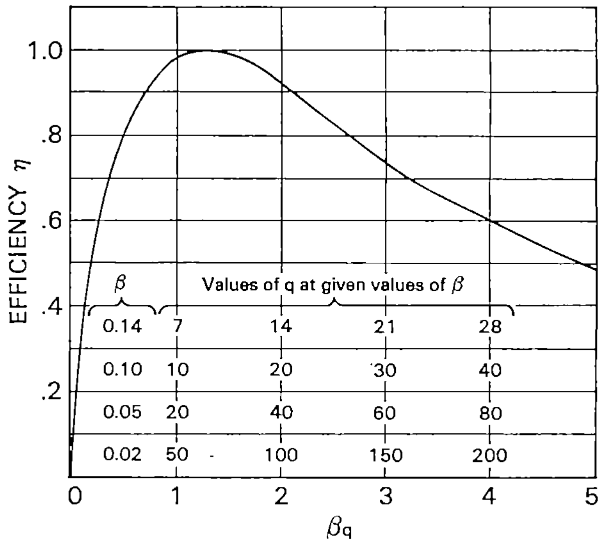


FIG. 5. The dependence of the efficiency on βq; for reference, the efficiency in function of the frame number for various β is also shown

or

$$q_{max} \simeq 1.26/\beta = 1.26P/D \tag{23}$$

For practical applications, the maximum is relatively flat, for $0.7 < \beta q < 2.1$ the efficiency is greater than 0.9. Fig. 5 shows also frame scales q for various values of β.

In applying this result to the NASA-DOD collection of Class B documents — open literature — one finds for $\beta = 0.02$ the most efficient microfiche size is obtained for q=63 frames. The present standardization is for q=60; therefore within practical limits the optimum is well achieved. Further significant advances in the state of the art in reduction (i.e., by a factor of 2) must be followed by a proportionate reduction in frames per microfiche. For the "older" microfiche sizes of 84 frames, the efficiency is 0.96. If European standardization of q=36 had been adopted, the efficiency would be 0.9. For type A documents, the present 60 frame size is uneconomical; for example, the efficiency for the AIAA collection is 0.29.

From equation (17) the optimum ratio of documents to microfiche is 0.72 (for $\beta q = 1.26$). Similarly, from equation (20) the optimum ratio of pages to frame becomes 0.57.

Documentary Relevance and Structural Hierarchy

JEAN M. PERREAULT[1]

Florida Atlantic University
Boca Raton, Florida

There is no writer today more closely identified with research into the idea of *relevance* as operative within the field of documentation than is Donald J. Hillman. It is not certain that he sees any real hope for arrival at any sort of technique capable of producing document-surrogates relevant to real queries — thus excepting those that satisfy queries derived from "source-documents." Yet, in his paper "The Notion of Relevance (I)" (1) he cites, presumably because it is relevant to his own discussion, a paper by B. C. Vickery, "The Structure of Retrieval Systems" (2). No classifier, surely — probably not even an indexer — would have thought of using any such conceptual indicator as "relevance" for Vickery's paper. If (and it is really conditional) we agree with the proponents of citation-indices, it must be that Vickery's paper is relevant to Hillman's topic. But why? Even more, *how* can a technique be devised capable of mechanically tracking down Vickery's paper as a result of the query, "What is available on relevance?"

Hillman puts the idea of relevance into question, but it seems to me that even so, no one would be unable to say quite a bit about the idea, or at least to decide whether "document Z is relevant to query A." The term itself has become fashionably popular since Cleverdon's test-results were phrased in such a way as to emphasize relevance and recall above all other possible criteria of the evaluation of retrieval tests(3). Yet it cannot be evaded that some things that someone thought (when he indexed a particular document) were relevant to certain topics were (from the querist's point of view) not so;[2] it is this disparity that lowers the relevance ratio. Such disparity is also called (in conformity with another fashion, that of "information theory") *noise*. Also, some things the querist would (presumably, if he were able to get at them) call relevant, but which the indexer did not, are missed, lowering the recall ratio.[3] This disparity could also be called (in conformity with the same fashion) *silence*.

But another factor enters, which most reference librarians have encountered many times: purely conceptual relevance (match, fit between query and document-orientation) is not enough for a patron in the real world, even if it may be good enough in a test situation. Even if all+only (conceptually) relevant documents are retrieved, they may not each be of use to the querist; formal and/or nominal considerations (4) may make even what is conceptually relevant inappropriate to a particular query.

One characteristic of relevance, then, as anyone could have seen, is that a judgment on it is a *value*-judgment.[4]

In both cases of disparity noted, leading either to *noise* or to *silence* — or in cases of formal or nominal inappropriateness — there is a sort of block involved which the temporal nature of storage and retrieval brings with it, since it is describable as "long-duration information-handling" or as a "delayed message-center." A situation could, of course, be imagined where the creator of surrogates was in possession of all the queries that were to be allowed, before analysis of the corpus; it would now be unnecessary to create surrogates, but instead only to check each document for relevance to each query. But even in this situation the querist, having delegated the relevance-judgments, could be disappointed by the results *if he too had access to the corpus.*

The future query, though, in the normal situation, cannot be anticipated in its own concreteness (as the past/present one can be), so the surrogate-creator attempts to indicate instead the orientation(s) *of the document.* "Concreteness" in this context is a bit of a technical term, signifying the philosophical concept that an entity is actual when it has received ("is concretized out of") all its formal and material perfections — or, that an essence or

[1] Head, Information Retrieval Division, and Associate Professor of Philosophy, Florida Atlantic University, Boca Raton, Florida.

[2] Or which were retrieved because of faulty syndetic trails.

[3] The fact that, in the "real world" of patrons and reference-librarians, the condition is that the querist *cannot* get at all the relevant documents unless the system reveals them to him, and that therefore the only legitimate technique for evaluation of performance is reading, by the querist, through the whole corpus to determine which (additional) relevant documents ought to have been supplied, and which supplied documents ought to have been omitted, effectively vitiates all critiques of the source-document-based "artificial" questions used by Cleverdon.

[4] *Cf.* the work cited in (3) for a statement that, while certainly putting relevance among values, tends to make it so subjective as to be quite ineffable: "Relevance cannot function as a criterion since there is no standard against which relevance judgments can be measured."

nature is made up of several intersecting or overlaid formalities. From this point of view, the *con*creteness of a document that could be *dis*cretely coded as A, B, C, D, could be exemplified (in particular by the punctuation) thus: A:([B+C](D)), or the like. The concreteness of the query is postulated as unlikely ever to match such a notation, but instead to be B(E)+F, or A_1:C(δ), or the like. There are quasi-intersections between individual (discrete) formalities in each of these configurations, such (perhaps) as to make the document relevant to the queries.[5]

The desideratum in mechanized retrieval is this, that the "perhaps" be validly removable; that, in other words, formal conformity be equivalent to material relevance (and hopefully, by the aid of formal and nominal elements in the document-surrogate, appropriateness as well); and, in still other words, that the only operation required of the mechanism be a clerical one. But, as noted above, the real problem is not "that" or "whether," but *"how."*

If, in our example, A_1 signifies a species of the genus A, what are the conditions under which the document A:([B+C](D)) would be relevant to the query A_1 (ignoring the rest of the original query-example for the moment)? It is simply this, stated as a binding convention:

> The concepts A_1, A_2, . . . A_n are so named because of the normality of their treatment together (= A). Where thus treated together, A is the indicator. Where treated in only partial togetherness, A is no longer totally appropriate, and must be omitted in favor of those of its species actually present.

For instance, if in the upper genus "dogs" there is a division by size, at least one middle genus will result as "miniature dogs" — besides other such middle genera. If the middle genus "miniature dogs" is itself divided into Pekinese, Chihuahua, *etc.*, it can itself be validly applied only when all the named species are present in the document. The general principle is thus:

> No genus may be pre-scribed as including a set of species without literary warrant for normality of such conjunction; no species may be pre-scribed as included in a genus except for converse reasons. Nor may any genus be in-scribed unless the document so inscribed treats the various species pre-scribed to that genus; nor, *mutatis mutandi,* any species.

Hence, if our example reads such that A is "miniature dogs" and A_1 is Pekinese, our convention indeed insures that A:([B+C](D)) is relevant to A_1. But, probably, other formalities as in our example —A_1:C(δ) — will be present in each query. Postulating that the document-surrogate can be translated as "influence of cold (B) and insufficient nutrition (C) as characteristic of high altitudes (D) upon miniature dogs (A)"; and postulating the query to be translatable as "influence of insufficient nutrition (C) as characteristic of abnormal environments (δ — the upper genus of D) upon Pekinese (A_1)," it can be seen that a purely clerical operation of the mechanism will produce the document in response to the query, and that its relevance is guaranteed by the convention.

Unfortunately, this solution extends no further than what can be called "explicit relevance." Such a case as was the starting point of this paper may not be included in it. But we do catch a glimpse of another characteristic of relevance, that a judgment on it is a *hierarchic* value-judgment. However, another look at what underlies the convention may furnish additional clues for a fuller solution. It has been proposed that classification is not necessarily hierarchical (5). I cannot agree that it is a legitimate use of the word "classification" to apply it to a mode of document-surrogation which does not have the characteristic "hierarchical." This presents the problem of an explicit distinction between (hierarchical) classification and those modes of document-surrogation which are non-classificatory. These other modes I would characterize as "indexing" — in any case, a clearer distinction has been needed for some time (6).

Indexing itself can be of at least two types, depending on whether it is controlled or uncontrolled in its vocabulary; this distinction could be put also as "whether its vocabulary is external or internal to the document(s)." If controlled, it *a*-scribes concepts to a document; if uncontrolled, it *de*-scribes the document by the words in it. Thus the use of controlled subject-headings,[6] just as much as the setting-up of a concordance, is indexing; they are both (-)scription of *elements of the documents.*

Classification, on the other hand, is not concerned with the detection of concepts or words within documents, at least not as a final purpose. It is concerned instead with *pre*-scription of positions within a systematic conceptual organization, and with *in*-scription of the documents to such positions and of such position-indications to documents and surrogates. Thus classification, too, is (-)scription, not of the elements of documents, but *of documents as wholes and of the corpus as a whole.* This may connote the marshalling function of traditional library cataloging/classification, or, on the other hand, the *concrete prescription* (recall concreteness as defined by the intersection of formalities) of articles in the *Revue Internationale de la Documentation, Referativnyi Zhurnal, etc.,* to complex UDC numbers.

The systematic-conceptual-organizational aspect of classification is also characteristic, to some extent of controlled thesaural subject-headings, where, if fully graphed out (7), the syndetic aspects would result in a systematic

[5] The difference between this situation (storage for future retrieval) and the hypothetical one outlined where there is no storage, is, besides the lack of futurity-based modes of effort, a wholly different attitude of analysis, depending on the possession or lack of the queries that are to be satisfied: what the proponents of natural-language whole-text "retrieval" are really striving for is removal of the "re-" by putting the computer in the place of the non-surrogate-creating analyst. Again, the lack of (at least some future) queries is what leads to the creation of surrogates, and is what forces analysis according to all that the analyst *can* have at hand: the document itself.

[6] In the sense that they are controlled, what Mooers calls 'descriptors' must be re-named 'ascriptors,' since if controlled they are *not* taken *from* ('de-') the document, but are ascribed *to* ('a-') it.

conceptual organization, however inadequate or difficult to follow through. It is not classification, though, in that it avoids going beyond the discrete level of indication of topical orientation — except insofar as the idea of roles and links has been adopted by indexers.

Diagrammatically, the new terminology would fit together as in Fig. 1.

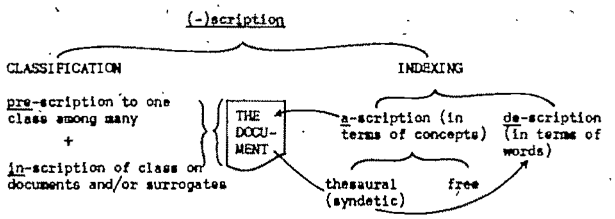

(-)scription
CLASSIFICATION
INDEXING
pre-scription to one class among many
+
in-scription of class on documents and/or surrogates
THE DOCU-MENT
a-scription (in terms of concepts)
de-scription (in terms of words)
thesaural (syndetic)
free

FIG. 1

[As a note to the idea of classification as necessarily hierarchical, I will not deny that, even within a hierarchical classification, classes do not stand in a hierarchical relation to *every* other class. Thus, in such a *schema* as in Fig. 2 the blocked-in classes are not *among themselves* hierarchically arranged. But, just insofar as these notations are conceptually filled and are used in a verbal system (if, then, α=fishes, D=felines, F_2=Pekinese), there are *implicit* (systematic [8]) relations between them which constitute hierarchicality *as soon as* the remaining terms are used (if x=animals, β=mammals, F=canines, etc.).]

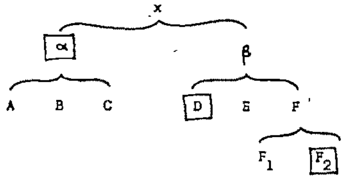


FIG. 2

Going back to the beginning, it can be said that relevance is a sort of double negative. I might very well be able to find a Biblical passage relevant to this paper, just as Hillman found a passage in Vickery; that is, without the thematic orientation of the cited bibliographic entity being such as to indicate the relevance. The Bible is further off, granted; but that is not really why I didn't go to it in search. . . . I haven't done so, not because it's not (possibly) relevant, nor because it's not particularly appropriate, but most of all because there is a fairly plentiful supply of more clearly relevant material nearer by — mostly because of the work of Hillman himself. Hillman, on the other hand, desirous of locating a relevant source by a documentalist (rather than more citations from Carnap, Goodman, or other philosophers — note the need for *appropriate* relevant documents), looked for a double negative: a document on documentation that was not relevant to relevance.

Again, though, how can such a process become purely clerical, in order to be fit for the delicate constitution of the computer? What good (in line with my own predilections) can hierarchy be for such a need?

Utilizing a hierarchicality that is *structural notationally* as well as conceptually, the mechanism can easily find the relevant document for the Pekinese query (the subscript "1" is analogous to a decimal structurality) — of course assuming the binding force of our convention. This can be reinforced all the more if the document-surrogate, in place of a merely conjunctive colon (= "mutual influence of some sort"), uses a more precise relational term (e.g., "is destroyed by"), and if the query is in such terms as (instead of its colon) "is injured by"; in a system where not only substantives but relatives are notated structural-hierarchically (9) such a query could likewise be satisfied by the document-surrogate mentioned, because of a subordinate relation between the query's relational term and that of the surrogate; besides helping prevent this document's becoming *noise* for a query where the relation is strongly variant from that mentioned, but the substantives remain the same.

The necessary factual condition that would now lead to the discovery of the Vickery document because of the insertion of a query phrased however Hillman did phrase it, is this: there is a particular system of document-surrogates organized conceptually into substantive classes and subclasses; capable of being mechanicoclerically manipulated because of the structurality of its notation; and capable of each part within it becoming modulated by any other part, in accordance with a relational *schema* similarly structural, hierarchical, and general-categoric — such that the terms of the query are related to the terms of the document-surrogate by pre-established paths.

Such a condition might, in this case, be met by classification "X," but not by "Y";[7] yet tomorrow there might be another query for which "Y" would produce far superior results. It is impossible, therefore, to say that any particular substantive organization (for, since the other characteristics should be made to obtain in *any* classification intended for use with the computer, *this* is what distinguishes one such system from another) is always best (LC? UDC? CC? *etc.*); from this many have validly but wrongly[8] concluded that classification — indeed, the creation of any kind of surrogate — is inadequate as the basis for mechanized retrieval, and that some means of automatic extrapolation from the full text of the author's own verbalization of his thoughts is necessary.[9] But it can be fairly conclusively shown by the testimony of those favorable to such a technique (10) that even such automatic means of creation of document-surrogates is far more successful in regard to classifying than in regard to indexing. And we must take our pick — there are no other available modes of conceptual bibliography.

[7] A substantive factor that would prevent retrieval of the relevant document for the Pekinese question would be, for instance, the non-inclusion of high altitudes (D) in abnormal environment (δ).

[8] As E. H. Gilson puts it, "in so far as logic is concerned, one may be faultlessly wrong as well as faultlessly right," *Being and Some Philosophers* (Toronto, Pontifical Institute of Mediaeval Studies, 1049), p. 100.

[9] *Cf.* footnote 4.

So our previous criteria still stand. We see that even a computer-prescribed classification will not necessarily always do for our querists what they would like (though, here as otherwhere, *silence* is golden in comparison to *noise:* if the querist doesn't know the answer already — for instance by having read the whole corpus prior to his query — not telling him anything will not reveal the defects of the system as much as will telling him something he *can* see is irrelevant). The question, then, is: Can there be one "perfect" computer-produced pre-scription/in-scription? Or, is it just like the "manual" classifications, a matter of some kind of taste or predilection which guides the creation of such a systematic conceptual organization? Does factor-analysis result in invariably reliable structures — or does associative mapping? Note at once the essential point: if one technique of automatic prescription is indeed invariant (so that queries can be constructed in accordance with it), there is surely no need for another technique unless the second results in advantages *not* found in the first — even if omitting some of the first's advantages.

The task ahead, then, is to classify classifications, both "manual" and automatic, (1) in accordance with their degree of satisfaction of our criteria, (2) in accordance with the substantive structural primordia upon which they are based, and (3) [which may be but another way of putting (2)] in accordance with their attitude toward and provision for relevances.

References

[*Cf.* in general Farradane, J. E. L. Classification and Mechanical Selection, *Proceedings, International Study Conference on Classification for Information Retrieval, Dorking, 1957* (London, Aslib): 65–69, 106–107. (1957).]

1. Hillman, D. J., The Notion of Relevance (I), *American Documentation*, **15**: 26–34 (1964).
2. Vickery, B. C., The Structure of Retrieval Systems, *Proceedings, International Conference on Scientific Information, Washington, 1958* (Washington, National Academy of Sciences), **2**: 1275–1289 (1959).
3. *Cf.*, for a much broader list of such criteria, *Summary, Study Conference on Evaluation of Document Searching Systems and Procedures, Washington, 1964* (Washington, National Science Foundation): 6–10 (1965).
4. Perreault, J. M., The Catalog and the Problems of Bibliography, sect. D.4, *Libri*, **15**: 287-339 (1965).
5. *Cf.* for instance Taulbee, O. E., New Mathematics for a New Problem, presented at the Conference on Electronic Information Handling, Pittsburgh (1964).
6. *Cf.* for instance Mills, A. J., Classification as an Indexing Device, Working Paper no. 16, presented at the International Study Conference on Classification Research, Elsinore (1964).
7. Perreault, J. M., The Conceptual Level in Bibliography, *Libri*, **15**: 302-310 (1965).
8. Scheele, M., "Significance and Problems of the Theory of Order," in *Punched-Card Methods in Research and Documentation, with Special Reference to Biology* (New York, Interscience), p. 114–127 (1961).
9. Perreault, J. M., Categories and Relators: A New Schema, *Revue Internationale de la Documentation*, **32**: 136–144 (1965).
10. Stevens, M. E., *Automatic Indexing: a State-of-the-Art Report*, National Bureau of Standards Monograph 91 Washington, U. S. Government Printing Office, p. 99 (1965).

Brief Communication

Mechanical Translation by Coordinate Indexing

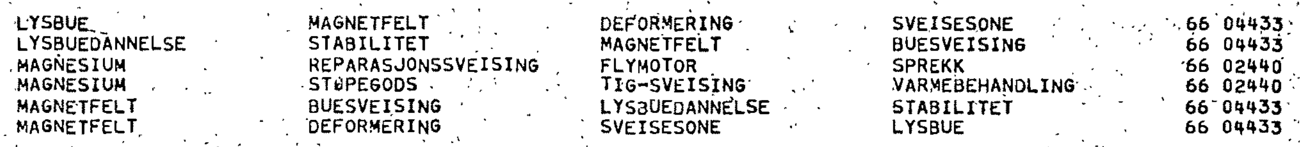
We certainly all realize that the language barrier is a serious hindrance in documentation work, and especially so for large-scale dissemination of information in science and technology. Maybe one day we shall live to see the automatic translation of natural languages come through as a working tool in documentation work, but so far we have to accept that semantic and syntactic problems involved in a full translation are still too difficult to be handled by documentation centers.

For this reason it might perhaps be of some interest to the readers of *American Documentation* to see some preliminary results from an effort to make a short-cut towards mechanical translation through the use of coordinate indexing. As described in an article in *Rev. Int. Doc.*, Vol. 32, No. 3, 1965, we are basing the information activities in our own organization on the use of a permutation index with precoordinated index terms.

A simple word-for-word translation involves, of course, no serious problems with a computer. With a precoordinated permutation index, however, the problems are indeed somewhat complicated, because the index terms have to be kept in their precoordination, and are finally being presented as an alphabetically arranged permutation index. In the UNIVAC 1107 which is used for the processing of our indexes, the programme now includes such automatic translations of the permutation index, through simultaneous use of bilingual vocabularies read onto magnetic tapes. With automatic translation for coordinate indexing, semantic and syntactic problems are no longer of serious importance. Semantics are taken care of through the controlled vocabulary or Thesaurus, and, apart from the use of index aids such as roles and links, there is no strong need for expressing syntaxes between the index terms to form a meaningful sentence.

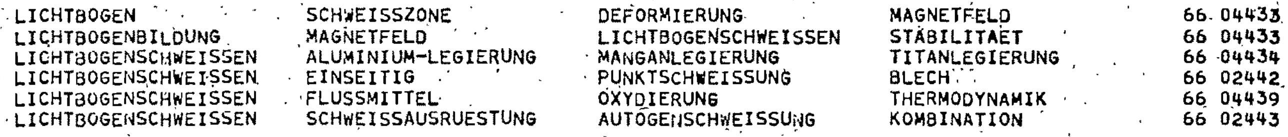
Problems have, however, been encountered in finding the exact translation of each of the index terms for the basic bilingual vocabularies, because of lack of congruence between the different languages. What we are aiming at in this case, however, is not an exact and lexically correct word translation, but merely a translation into a concept which does convey to the reader the same basic idea. For this reason the bilingual vocabularies used are nonreversible, being used one way only.

The illustrations show some samples from the Norwegian, English, German, and Russian versions of this mechanically translated permutation index for a collection of welding



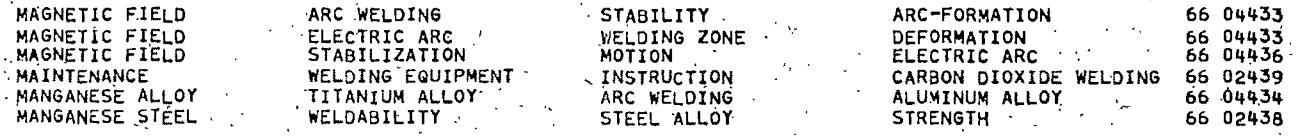
```
LYSBUE              MAGNETFELT             DEFORMERING            SVEISESONE             66 04433
LYSBUEDANNELSE      STABILITET             MAGNETFELT             BUESVEISING            66 04433
MAGNESIUM           REPARASJONSSVEISING    FLYMOTOR               SPREKK                 66 02440
MAGNESIUM           STØPEGODS              TIG-SVEISING           VARMEBEHANDLING        66 02440
MAGNETFELT          BUESVEISING            LYSBUEDANNELSE         STABILITET             66 04433
MAGNETFELT          DEFORMERING            SVEISESONE             LYSBUE                 66 04433
```

FIG. 1. Norwegian



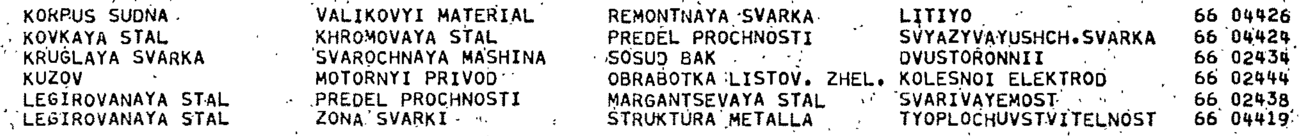
```
LICHTBOGEN              SCHWEISSZONE           DEFORMIERUNG            MAGNETFELD             66 04433
LICHTBOGENBILDUNG       MAGNETFELD             LICHTBOGENSCHWEISSEN    STABILITAET            66 04433
LICHTBOGENSCHWEISSEN    ALUMINIUM-LEGIERUNG    MANGANLEGIERUNG         TITANLEGIERUNG         66 04434
LICHTBOGENSCHWEISSEN    EINSEITIG              PUNKTSCHWEISSUNG        BLECH                  66 02442
LICHTBOGENSCHWEISSEN    FLUSSMITTEL            OXYDIERUNG              THERMODYNAMIK          66 04439
LICHTBOGENSCHWEISSEN    SCHWEISSAUSRUESTUNG    AUTOGENSCHWEISSUNG      KOMBINATION            66 02443
```

FIG. 2. German



```
MAGNETIC FIELD      ARC WELDING            STABILITY              ARC-FORMATION            66 04433
MAGNETIC FIELD      ELECTRIC ARC           WELDING ZONE           DEFORMATION              66 04433
MAGNETIC FIELD      STABILIZATION          MOTION                 ELECTRIC ARC             66 04436
MAINTENANCE         WELDING EQUIPMENT      INSTRUCTION            CARBON DIOXIDE WELDING   66 02439
MANGANESE ALLOY     TITANIUM ALLOY         ARC WELDING            ALUMINUM ALLOY           66 04434
MANGANESE STEEL     WELDABILITY            STEEL ALLOY            STRENGTH                 66 02438
```

FIG. 3. English



```
KORPUS SUDNA        VALIKOVYI MATERIAL     REMONTNAYA SVARKA      LITIYO                   66 04426
KOVKAYA STAL        KHROMOVAYA STAL        PREDEL PROCHNOSTI      SVYAZYVAYUSHCH.SVARKA    66 04424
KRUGLAYA SVARKA     SVAROCHNAYA MASHINA    SOSUD BAK              DVUSTORONNII             66 02434
KUZOV               MOTORNYI PRIVOD        OBRABOTKA LISTOV. ZHEL. KOLESNOI ELEKTROD       66 02444
LEGIROVANAYA STAL   PREDEL PROCHNOSTI      MARGANTSEVAYA STAL     SVARIVAYEMOST            66 02438
LEGIROVANAYA STAL   ZONA SVARKI            STRUKTURA METALLA      TYOPLOCHUVSTVITELNOST    66 04419
```

FIG. 4. Russian

documents. As one of the first results these experiments have clearly demonstrated the importance of an exact indexing from the side of the indexer, and full stringency in the selection of index terms. The Russian version of the index can now be printed also with cyrillics through the use of Siemens' teletypewriter T type 100. In this case the bilingual vocabulary contains the necessary code for cyrillic letters, and output from the computer is a 5-channel paper tape to operate the telex.

W. Holst, Adm. Director
Norwegian Industries Development Association
Forskningsveien 1
Oslo 3, Norway

Users Versus Documents

For the first time in these pages (or in any pages as far as I know), Mantell[1] has provided reasonable and justifiable estimates of the amount of technical papers being produced. It would appear, from these estimates, that many of us have been crying "wolf." The "information explosion" is seen to be a rather small bang. Lest anyone be lulled into a false sense of insecurity by the loss of our most quoted "statistic," let me reiterate a point often made before,[2,3] namely, that our problems stem not simply from documents or users, but rather from the interactions between them. Thus the size (or rate of increase) of *our* problem as distinguished, say, from that of the publishers or universities must be reckoned in terms of the *product* of users and documents.

The accompanying graph (Fig. 1), using Mantell's (i.e., NSF's) figures (and the most conservative for 1970), shows the anticipated growth of scientists and engineers (S + E), i.e.. users (U), documents (D), and their interaction (UD). The latter has been normalized to fit the scale of the graph.

The information specialists job, after all, is to *match* users to documents and vice versa. UD represents all possible matches that one might examine. In a linearly organized file on a serial access machine[4] or in most SDI systems,[5,6] all of UD must be examined. Humans, luckily, have content-addressable memories. Depending on the efficiency of the matching algorithm that they use, the proportion of UD that must be examined may be significantly reduced. This is the best argument I know for the need for random-access, content-addressable memory for information retrieval systems.

References

1. Mantell, L. H. 1966. On Laws of Special Abilities and the Production of Scientific Literature. *Am. Doc.*, **17** (1): 8–16.
2. Savage, T. R. 1963. Is Yehoshua Bar-Hillel Approaching a Crisis? (Letter to the Editor.) *Am. Doc.*, **14** (4): 331.
3. Hensley, C. B., et al. 1962. Selective Dissemination of Information — A New Approach to Effective Communication. *IRE Transactions on Engineering Management* EM-9(2): 55–65.
4. Swid, R. E. 1963. Linear vs. Inverted File Searching on Serial Access Machines. Short Papers — Part 2, ADI Annual Meeting. *Automation and Scientific Communication*, 169–170.
5. Hensley, C. B., et al. *op. cit.*
6. Brandenberg, W., et al. 1961. Selective Dissemination of Information, SDI 2 System. Advanced Systems Development Division, IBM, Yorktown Heights, N. Y.

T. R. Savage
Datatrol Corporation
Rockville, Maryland

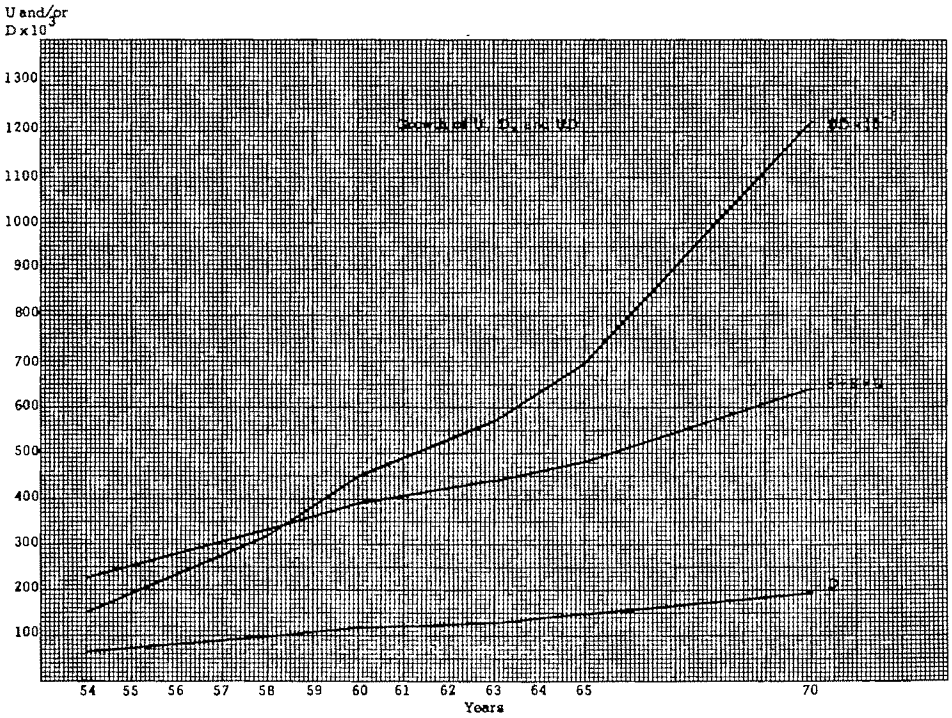


Fig. 1. Growth of Users, Documents, and UD

Reversible Punctuation Russian Transliteration

In processing Russian scientific papers, we need to reproduce the original spelling of English names which have been represented by Russian spellings. At present, we can't be sure that our English respelling is accurate when, as periodically occurs, our only clue is the Russian spelling of a name originally in English.

There are reasonably satisfactory Russian to English transliteration systems. There is no accurate English to Russian counterpart. What does exist is a subjective approach based on Russian assumptions as to how the English name is pronounced, which is inadequate even when the pronunciation assumptions are correct.

Since there are 7 English letter concepts (of the 26), which have no single transliteration equivalent among the 33 Russian letter concepts, some substantial innovation is clearly required if we are to achieve reversible processing of the 40 different letter-concepts in the combined Russian-English alphabet.

While there are some sounds in English not found in Russian, and some sounds in Russian not found in English, the real transliteration problem arises from the differing spirits of the two alphabets. Russian is not always phonetically unambiguous, but its users try to be. Users of English are content to rely on groups of letters and associations of sounds for hints as to pronunciation.

The very valuable discussion of Russian transliteration in *Science* (1–3) made some of these points while leaving others insufficiently clear and explicit.

As other comments have suggested more or less explicitly (4–6), the real problem is getting the Soviet Russians to adopt a transliteration of the English or Latin alphabet into Russian which, in general, will be reversible in that it permits the original English or Latin alphabet spelling of a name to be reproduced automatically. To solve our problem, we need to persuade the Russians to abandon a phonetic transcription approach for indicating English-type alphabet names in scientific contexts.

Of the 40 different letter concepts in the two languages, 19 (see Fig. 1, B and C) already have single letter symbol equivalents in both languages for transliteration from Russian into English. Two Russian letter-concepts, the hard sign (or, in Bulgarian, the neutral vowel), and the soft sign, which gives a y sound to preceding consonants and following vowels, are already customarily represented in English by the punctuation symbols " (quotation mark) and ' (apostrophe). (Strictly speaking, the quotation mark or apostrophe used as punctuation should be separated when included in transliterated material by one space from what it modifies, and by two spaces from words not modified. Note that Cyrillic [Russian-type alphabet] apostrophe is transliterated into Russian as N° followed by soft sign, and into English as ;' [semicolon apostrophe] [cf. 7].)

Three Russian letter-concepts are customarily represented by the English two-letter combinations zh, kh, and sh, without ligatures or diacritical marks or accents. (For complete unambiguity in automatic machine transliteration, z:h, k:h, and s:h might be desirable.)

Before we consider the other 16 Russian-English letter-concepts, we must comment rather firmly on such remarks as "We may simply ask our printer to supply the non-available letter [i with a cup over it]" (2, pp. 485–486). The controlling factor is not what we can persuade some printers to do at extra expense, but what the average English or Russian typewriter will do, as is.

For at least 3 Russian-English letter-concepts(t:s, sh:ch, and the Russian transliteration of English x as k:s), we need a ligature symbol to indicate that two or more symbols in one language represent a single symbol in the other. The dash (hyphen) or absence of a mid-letter-level dot are inadvisable for this. Since the colon between two letters visually suggests letter:ligature, and since the colon is never followed directly by a letter when used as punctuation, it is used as the general sign of ligature.

In general, we use preceding , (comma) for the soft sign or any cup or v mark over a letter, and as a sign of aspiration; two dots (..) preceding for a diaeresis or umlaut over a letter; and following slant or virgule (/) to indicate that the original letter symbol had an extender.

We use the Russian equivalent of ,i/ to transliterate English y into Russian and yh and eh for the English transliterations of the Russian "hard" i and "hard" e letter-concepts. We can then adopt the current Russian to English transliteration usage of yu and ya for the last two Russian letters.

A general transliteration use of the preceding English semicolon and Russian N° (number) sign is to indicate that the original symbol for a transliterated Cyrillic letter concept is visually identical with a letter in the English alphabet. The combination ;ch to indicate the Russian letter-concept pronounced t:sh is a special use of the semicolon in English.

For complete unambiguity in automatic machine transliteration, y:h, e:h, y:u, and y:a can be used.

The use of the Russian equivalent of ,v to transliterate English w is by analogy with the Belorussian letter ,u. We use a ligated doubled letter as in the Russian equivalent of k:k for English q, generally to represent the class of letters to which q originally belonged. In some Cyrillic alphabets, a form which would be transliterated as the Russian equivalent of kh/ is used to represent a characteristic sound of English h.

The phonetically ambiguous English letter-concept c, which we transliterate by the Russian equivalent of t:s/ , shares one of its several sound values with the unambiguous Russian letter-concept t:s. It is hoped that the "second order extender" aspect of the Russian equivalent of t:s/ may suggest the intrinsic ambiguity of the English letter-concept c to Russian users.

In transliterating j, however, it was felt that the Russian equivalents of d:zh/ or zh/ or '/ would fail to convey the intrinsic ambiguity of the letter-concept j, and the Russian equivalent of d:z/ was consequently used to transliterate English j.

The 14 modifications in present Russian to English transliteration practice (including yh and eh, but not the current " and ') are summarized in Fig. 1, A.

A good system should entail as few changes as possible in the existing alphabetization of Russian names and titles in English. The proposed punctuation transliteration system requires only two changes (eh and yh from e and from y or i) in the English alphabetization of transliterated Russian names now used by *Biological Abstracts,* BSI and ASC Z39; *Chemical Abstracts, Applied Mechanics Reviews,* and *Science Abstracts.* The British Museum would also use " and penultimate i instead of omitting a transliteration for these letters, and would use yh for ui. The USBGN would use ,i instead of y and initial e instead of initial ye (8–14). The Library of Congress now differs from all these systems in using iu and ia for yu and ya (15).

For some applications, it is desirable to assign a definite alphabetical order position to most of the punctuation and Arabic numeral symbols. This is done in Fig. 1.

The lack of a transliteration system guaranteeing the recovery of the exact original spelling of a name after it has passed through a letter-substitution "mill" (16) must also cause Soviet citizens inconvenience in connection with non-Russian Soviet, and other Cyrillic names entering Russian directly, and Latin alphabet Slavic names entering Russian via English.

Outside the Soviet Union, the Serbs and Macedonians (Yugoslavia), the Bulgarians, and the non-Slavic Mongolians use modified Cyrillic alphabets. Inside the Soviet Union, modified Cyrillic alphabets are used by the Ukrainians and Belorussians, and by about 52 non-Slavic Soviet nationalities. The Cyrillic alphabets other than Russian include about 41 additional Cyrillic letter-concepts, 14 duplicates which now have to be transliterated separately if the original text is to be reproduced visually, and 4 ambiguous symbols at least one use of which should be eliminated, but which now would have to be transliterated separately. (Fig. 1, D:) (17–21.)

The East Armenians in the Soviet Union (and the West Armenians elsewhere) have an alphabet of their own, as do the Soviet Georgians, and those Jews in the Soviet Union

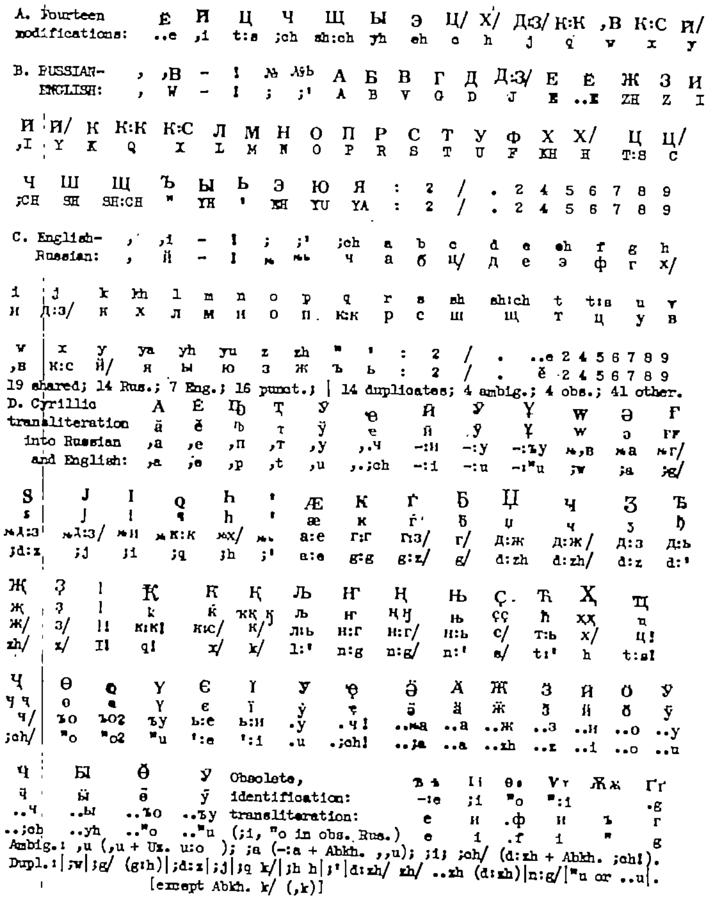


FIG. 1. Cyrillic transliteration into Russian and English

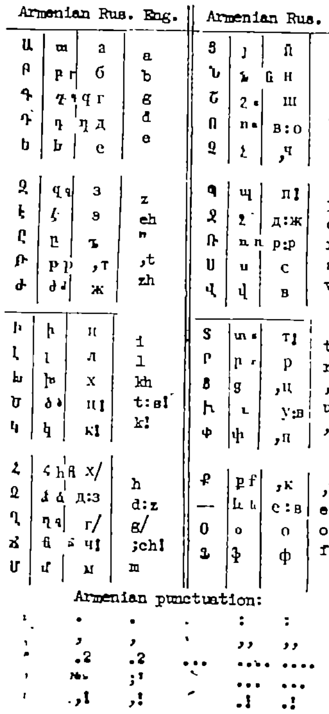
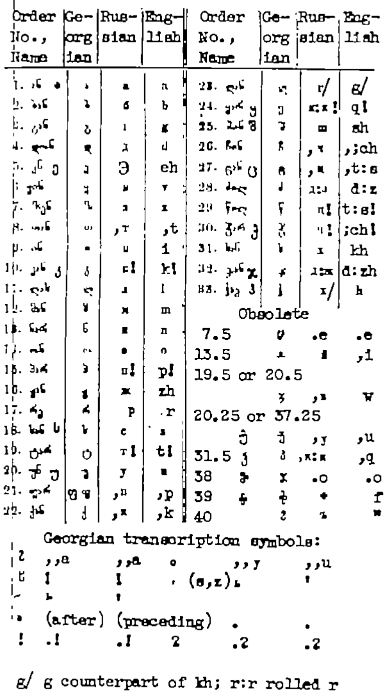


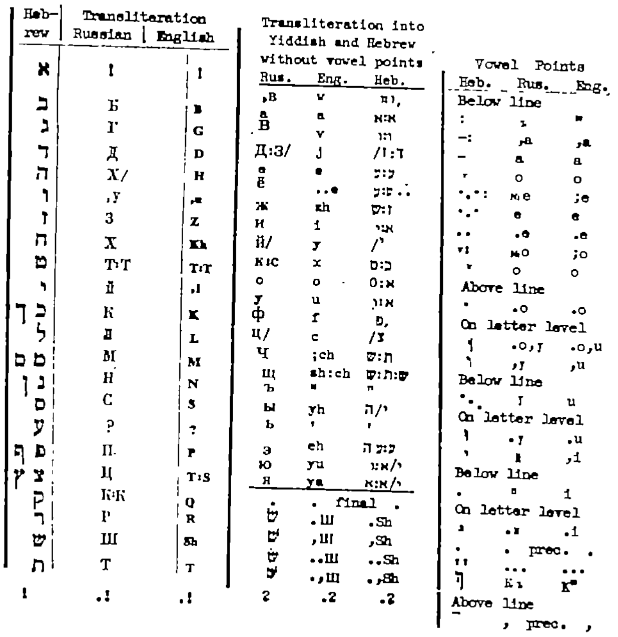
FIG. 2. Armenian transliteration into Russian and English



g/ g counterpart of kh; r:r rolled r

,: aspirate; :! glottalic abruptive.

FIG. 3. Georgian transliteration into Russian and English



Note: Two Russian letter-symbols and letter concepts are from the Hebrew alphabet:

T:S Ц צ SH Ш ש

FIG. 4. Yiddish and Hebrew transliteration into Russian and English. Russian and English transliteration into unpointed Yiddish-Hebrew. (Note: The Yiddish-Hebrew alphabet is read from right to left.)

who speak Yiddish and use the Hebrew alphabet for it. Armenian, Georgian, and Yiddish written in the Hebrew alphabet share some letter concepts with Russian, and also have some letter concepts which are not found in Russian.

We use preceding , (comma) to indicate aspiration, and a following exclamation point (!) to indicate a glotallic abruptive, and can then easily transliterate Armenian and Georgian into English or Russian (Figs. 2 and 3, notes 22–29).

(The West Armenians outside the Soviet Union, who use the same complete Armenian alphabet, do not pronounce the glotallic abruptives or the aspirates as such, and make 15 changes in the sounds of the letters. In West Armenian, the letter concepts which are aspirated in East Armenian [transliterated with a preceding , (comma)] lose their aspiration without interchange, and, aside from the loss of the glotallic abruptive pronunciation, there is interchange of the root phonetic values of the letters of the b and p!, g and k!, d and t!, d:z and t:s and d:zh and ;ch! pairs. The suggested transliteration system is based on the East Armenian used in the Soviet Union, but should suffice to transliterate any Armenian text, regardless of how the writer would pronounce it.)

For the Hebrew-Yiddish alphabet we need special symbols for the letter aleph (!, exclamation point, not directly related in this use to its use to represent glotallic abruptives), for the letter ayin (?, question mark), for the letter teth (t:t, doubled with ligature), and for the letter vav (,u) (30–33, 17 pp. 84–85, and 19, p. 27) (Fig. 4).

The punctuation system permits transliterating French acute accents (by preceding period), circumflex accents (by preceding /: slant colon), French diaeresis or German umlaut (by preceding double period), cedilla or extra extender on letter (by following slant [/], Spanish tilde (by :' [colon apostrophe] following letter), and French accent grave (by preceding semicolon [;] into English or by preceding N° [number] sign into Russian).

The Latin alphabet Slavic languages, related to Russian and of particular interest to the Russians, average six extra consonants, each differentiated by diacritical marks.

To transliterate the 11 Serbocroatian variants we need d:' (-:d), zh (,z), ;j (j), l:' (lj), n:' (nj), t:' (.c), kh (h), t:s (c), ;ch (,c), d:zh (d,z), and sh (,s).

To transliterate Czech we need .a ,c .e ,e .i ,n .o ,r ,s .u (acute accent initially, circle above letter medially and finally) .y and ,z. In Slovak, /:o can be used to indicate o circumflex. In both Czech and Slovak, ,D and d:' and ,T and t:' are variant forms of the same letter. Slovak also has ..a .l L:' and l:' .r and .u (u acute medial).

In Polish, we need to transliterate z with an acute accent as ,z, and to use z:sz for z with a dot over it, the z or zh analog of "hard" sz and cz. Polish transliteration then requires a/ .c e/ l/ .n .o .s .z (z with an acute accent) and z:sz (z with a dot over it).

With :' used to indicate a tilde-like sign over the original letter, -: used to indicate a horizontal line over or through a letter, and the standard usage of preceding .. for umlaut, preceding , for cup or v over a letter, preceding . for acute accent, and following / for an extra extender, the extra letters in the Soviet non-Slavic Latin alphabet languages, including Latvian, Lithuanian, and Estonian, can be transliterated as ..a -:a a/ ,c .e -:e e/ g/ -:i i/ k/ l/ n/ ..o o:i' ,s ..u ū/ and ,z (34–35, 17 pp. 75–79).

The punctuation transliteration system appears to be an adequate solution to the problem of reversible English-Russian-English transliteration of the names of scientists that would obviously be as useful to the Russians as to ourselves.

References[1]

1. Razran, G. 1959. Transliteration of Russian. *Science,* **129:** 1111–1113.
2. Hamp, E. P., A. C. Faberge, M. B. London, I. D. London, D. T. Ray, and G. Razran. 1959. Russian-English Transliteration. *Science,* **130:** 482–488. (Four separate comments and a reply.)
3. Susich, G., and G. Razran. 1960. Russian Transliteration. *Science,* **131:** 324.
4. Reformatski,i, A. A. 1960. Transliteration of Russian Texts by Latin Letters (in Russian). *Voprosyh Yazyhkoznaniya* (Problems of Philology), **9** (5): 96–103. (Folding table p. 102a.)
5. Kuznet:sova, V. I. 1960. *Foneticheskie osnovyh peredachi angli,iskikh imen sobstvennyhkh na russkom yazyhke.* (Phonetic Fundamentals of Rendering English Personal Names into Russian.) 120 pp. State Teaching-Pedagogic Press of the Min. of Education RSFSR, Leningrad Division, Leningrad.
6. Leont'ev, A. A. 1963. On the Spelling of Foreign Names (in Russian). *Akademiya Nauk SSSR. Institut Russkogo Yazyhka. Voprosyh Kul'turyh Rechi.* (Acad. Sci. USSR. Inst. Russ. Language. Problems of Speech Culture), **4:** 154–156.
7. Berezin, B. I. 1965. *Samou;chitel' Mashinopisi.* (Typewriting Self-taught.) 160 pp.+2 p. loose chart. "Kniga" ("Book") Press, Moscow. Second Edition, pp. 5–6.
8. Biological Abstracts. 1961. *Instructions on the Translation of Russian Scientific Text for Publication in Biological Abstracts.* 32 pp. Second Edition, *Biol. Abstr.,* Philadelphia, p. 8.
9. British Museum. 1936. *Rules for Compiling the Catalogues of Printed Books, Maps and Music in the British Museum.* 68 pp. British Museum, London, pp. 52–55.
10. British Standards Institution. 1958. *British Standard 2979:1958. Transliteration of Cyrillic and Greek Characters.* 23 pp. British Standards Institution, London, pp. 8–9.
11. Orne, J. 1963. Transliteration of Modern Russian. *Libr. J.,* **88:** 4157–4160. (ASC Z39 table on p. 4159.)
12. Chemical Abstracts Service. 1964. *Directions For Abstractors.* 76 pp. Chemical Abstracts Service of the Am. Chem. Soc., Columbus, Ohio, p. 4–2.
13. U. S. Board on Geographic Names. 1944. In U. S. Government Printing Office 1959 *Style Manual.* Revised Edition. 496 pp. Washington, D. C., pp. 474–476.
14. Applied Mechanics Reviews and Science Abstracts. In *Mathematical Reviews.* 1964. **28:** 1248 (Table: "Transliteration of Russian".)
15. Library of Congress. 1945. *Transliteration—Slavic. The alphabets of "Cyrillic" origin.* L. of C. Cat. rules (Suppl.) [r45k410] Rule 10—Rev. Jan. 12, 1945. Cards 1–4. Library of Congress, Washington, D. C.
16. Garfield, E. 1964. "Science Citation Index"—A New Dimension in Indexing. *Science.* **144:** 649–654. (p. 653).
17. Gilyarevskii, R. S., and V. S. Grivnin, 1964. *Opredelitel' yazyhkov mira po pis'mennostyam* (Manual for the Identification of the Languages of the World on the Basis of Written Specimens). 375 pp. Acad. Sci. USSR, Inst. of the Peoples of Asia, All-Union State Library of Foreign Literature, "Nauka" ("Science") Press, Moscow. Third Edition, pp. 299–300; 84–85; 75–79.
18. *Bol'shaya Sovetskaya Ehnt:siklopediya* (Large Soviet Encyclopedia). 1955. Writing (in Russian). **33:** 99–109. Second Edition. Moscow.
19. Musaev, K. M. 1965. *Aljavityh yazyhkov narodov SSSR* (Alphabets of the Languages of the Peoples of the USSR). 87 pp. "Nauka" ("Science") Press, Moscow. Tables, pp. 26 . . . 69.
20. Genko, A. N. 1957. Phonetic Interrelationships of the Abkhazian and Abazinian Languages (in Russian). *Akademiya Nauk Gruzinsko,i SSR. Abkhazski,i Institut Yazyhka, Literaturyh i Istorii. Trudyh.* (Acad. Sci. Georgian SSR. Abkhazian Inst. Language, Literature, and History. Transactions.) 28: 177–225.
21. Bgazhba, Kh. S. 1959. Notes on the History of Writing in Abkhazia (in Russian). *Ibid.* **30:** 245–290. Tables on pp. 287–290.
22. Fairbanks, G. H., and E. W. Stevick. 1958. *Spoken*

[1] The reversible punctuation transliteration system is used where applicable by way of illustration.

East Armenian. 403 pp. American Council of Learned Societies, N. Y., p. 74.

23. FAIRBANKS, G. H. 1958. *Spoken West Armenian.* 204 pp. American Council of Learned Societies, N. Y., p. 79.

24. VARTAPETYAN, N. A. 1961. *Spravochnik po russko,i transkript:sii armyanskikh imen* (Handbook on the Russian Transcription of Armenian Personal and Geographic Names). 131 pp. Armenian State Press, Erevan, pp. 10–54.

25. MOVSESSIAN, P. L. 1959. *Armenische Grammatik, West-, Ost- und Altarmenisch* (Armenian Grammar, West-, East-, and Old Armenian). 423 pp. Mechistaristen-Buchdruckerei, Wien, pp. 10–17.

26. MARR, N. 1926. *Posobie dlya izucheniya zhivogo gruzinskogo yazyhka* (Manual for the Study of the Living Georgian Language). Issue I. 99 pp. A. S. Enukidze Leningrad Inst. Living Eastern Languages, Leningrad, p. 1.

27. RUDENKO, B. T. 1940. *Grammatika Gruzinskogo Vazyhka* (Grammar of the Georgian Language). 275 pp. Acad. Sci. USSR, Inst. for the Study of the East (Vostokovedenie), Transactions, No. 32, Press of the Acad. Sci. USSR, Moscow-Leningrad, pp. 19–27.

28. TSCHENKELI, K. 1958. *Einf..uhrung in die Georgische Sprache* (Introduction to the Georgian Language). 2 vols. [1. LXIV + 628 pp.; 2. X + 614 pp.]. Amirani, Z..urich. Vol. 1, XLIII–XLV.

29. AKHVLEDIANI, G. S. 1959. On Some Fundamental Problems of the Phonetics of the Georgian Language (in Russian), pp. 39–67 (with English summary pp. 65–67), *in Tbilisski, i Gosudarstvennyh, i Universitet im. I. V. Stalina. Trudyh Kafedryh Obsh:chego Yazyhkovedeniya 3. Foneti;cheski, i Sbornik 1.* (I. V. Stalin Tiflis State University, Transactions of the Dept. of General Philology 3. Phonetics Symposium 1.) Ed. V. A. Artemov and S. M. Zhgenti. Tiflis.

30. SHAPIRO, F. L., and B. M. GRANDE. 1963. *Ivrit-Russki, i Slovar' Mi.L.o,uN ?iV"Ri,i-R,uSi,i* (Hebrew-Russian Dictionary). 766 pp. State Foreign and Nationality Dictionary Publishing House, Moscow, pp. 667–668.

31. KAUSHANSKI,I, M. M. 1935. *Inostrannyh, i Nabor* (Foreign Type-setting). 218 pp. State Light Industry Press, Moscow-Leningrad, pp. 180–189.

32. MORAG, S. 1962. *The Vocalization Systems of Arabic, Hebrew and Aramaic.* 85 pp. Mouton, 'S-Gravenhage, pp. 17–29, folding tables 78a and 79a.

33. *Bol'shaya Sovetskaya Ehnt:siklopediya.* 1957. Articles on the Russian letters T:S and SH (in Russian). **46**: 431 and **47**: 490. Second Edition. Moscow.

34. DE BRAY, R. G. A. 1951. *Guide to the Slavonic Languages.* 797 pp. Dent, London, pp. 73–74, 132–133, 196–197, 247, 316, 369–370, 439–440, 516–517, and 593–594.

35. Polski Komitet Normalizacyjny (Polish Committee on Norms). 1959–1960. *PN-59 N-021201 Przepisy bibliograficzne. Transliteracja alfabet.ow cyrylickich* (PN-59 N-021201 Bibliographical Rules. Transliteration of Cyrillic Alphabets.) 4 pp. [Warsaw.]

DAVID FRANKLIN
1601 56th Street
Brooklyn, N. Y. 11204

Simulation of Boolean Logic Constraints Through the Use of Term Weights

The evolution described below of one aspect of the NASA Scientific and Technical Information Facility's machine search system may be of general interest to the documentation profession.

The Facility began operations in early 1962. The literature search service, or "demand bibliography" service, as it was then termed, was initially a very modest endeavor for the simple reason that the data base upon which to search had yet to be built. The first search programs concentrated on the well-known Boolean logic capabilities in the searching of inverted term files on magnetic tapes. This was consistent with the contractor's (Documentation Incorporated) prior R&D experience with so-called "Uniterms" and coordinate indexing systems.

A major change was effected, beginning in January 1965, to a serial or linear type of file organization. The reasons for this change were many and varied and need not concern us in any detail here. They involved, primarily, efficiencies in the file maintenance and update procedures and in the journal index preparation procedures. Also, it was becoming imperative to be able to search the file on a variety of nonsubject, administrative categories of information. At the time of this change, additional capabilities were built into the new "linear" search system. To supplement the basic Boolean capability, we now, among other things, made available to ourselves the following strategies that were well known in the state-of-the-art: (1) a weighting technique, (2) a "root" searching technique, and (3) a system of nonsubject "limits."

The weighting technique permits the assignment of arbitrary weight values to search terms and the specification of a minimum weight which any document must achieve in order to become a "hit."

"Root" searching permits queries on any desired generic level of various entities, e.g., all contracts with the prefix NAS8-; all report numbers with the prefix RAE-; all authors with names beginning CAR-. It may soon be extended to index terms, as in all terms beginning "PNEUMO," etc.

The system of "limits" permits the specification of various additional constraints on a search other than those involving subject index terms. Nearly all the standard descriptive cataloging elements fall within this system.

Each of these new capabilities has seen a great deal of use. The weighting technique, however, has particularly caught the interest of the searching staff and has resulted in some far-reaching developments.

For instance, it is apparent that document weight becomes a way of ranking search output in order of relevance. Probably the first use that weights were put to within the Facility was not to limit the output — the Boolean equation did this — but to *arrange* it for either the user or the analyst or perhaps both. This became extremely valuable in an environment where search output received a human edit before it was released. Arbitrary weight levels could be set by the analyst above which relevance to the question was assumed and below which his editorial effort was concentrated.

It also became apparent that the weighting technique could, by itself in some situations, achieve exactly the same results as a Boolean equation; cleverly assigned weights could *simulate* such an equation. For example, the equation (1) $A(B + C + D) = \text{Answer}$, can be completely bypassed through the following weight assignments: $A = 3$, $B = 1$, $C = 1$, $D = 1$; Weight Limit = 4. This becomes very useful to know, for the calculation of weights was a much faster computer process than the solving of a Boolean equation, and the substitution could lead to significant computer time savings. Other common types of substitutions were the following:

(2) $A + B + C + D$	$A = 1, B = 1, C = 1, D = 1$ Weight Limit = 1
(3) $A \cdot B \cdot C \cdot D$	$A = 1, B = 1, C = 1, D = 1$ Weight Limit = 4
(4) $A + (B \cdot C \cdot D)$	$A = 3. B = 1. C = 1, D = 1$ Weight Limit = 3
(5) $(A + B) + (C \cdot D)$	$A = 2, B = 2. C = 1, D = 1$ Weight Limit = 2
(6) $(A + B) \cdot (C \cdot D)$	$A = 1, B = 1, C = 2, D = 2$ Weight Limit = 5

Various rules of thumb can easily be developed, and were, for the proper assignment of weights in more complex situations of the above basic types. However, no mathematical formalization was ever attempted.

It was soon realized that though term weighting had its advantages, nevertheless there were some equations that could not be reduced in this way. Two of the most basic are the following:

(7) $(A + B) \cdot (C + D)$

(8) $(A \cdot B) + (C \cdot D)$

The above equations cannot be simulated through any assignment to their terms of positive or negative weights, in conjunction with a weight limit. This can be proved by fairly simple algebraic techniques which will not be gone into here.

Continuing examination of the recalcitrant situations led to the development of a special "Group Weight" system for processing them. Essentially this involves "multiplying out" the equation, identifying its sections or groups, and assigning weights and weight limits for each section. Equation (7) thus becomes the redundant (7A) $A(C + D) + B(C + D)$ and weights may be assigned as follows:

Group A: $A(C + D)$	$A = 3, C = 1, D = 1$; Weight Limit $= 4$
Group B: $B(C + D)$	$B = 3, C = 1, D = 1$; Weight Limit $= 4$

The search program is now in the process of being changed to permit this technique. Logical equations will be made an optional, not a mandatory, feature of a search question. All types of logical equations may then be converted solely to a system of term weights and weight limits. Tests have been run comparing search times for ten problems coded by equation against the same ten coded with weights; both sets being run on our IBM-1410 search system against the same single reel of the data base. Results indicate that there is a 4 to 1 time advantage to running in the weight or arithmetic mode. However, it is clear that complicated equations can be both difficult and laborious to code. The next step is therefore obvious. In those cases where weights would be used mainly to simulate Boolean logic for the sake of processing speed, there is no reason that the program should not accept the equation and calculate its own weight assignments. This is now being evaluated.

It is thought that this particular case history in the use of weights may be of interest because of the widespread current use of weights in machine search systems. Several systems seem to be dropping the Boolean capability per se altogether in favor of weights. The two are generally spoken of in these situations as disparate entities. It is not that simple. The closeness of the relationship is shown by the fact that the weighting technique can be made to simulate Boolean logic. However, in doing so, the weighting technique can easily become too difficult for convenient human use. On the other hand, the logical equation is perhaps the most unambiguous and easily comprehensible way a search question with a complex relationship of terms can be organized and displayed. Our own solution is to keep both strategies in order to take advantage of the *unique* capabilities that each has to offer. At the same time, we are attempting to take advantage of the newly realized (at least as far as we are concerned) relationship between the two systems by utilizing the fast weight calculation process as a technique for internal computer solving of a logical equation.

W. T. Brandhorst
Assistant Director for Operations
NASA Scientific and Technical Information Facility
and
Documentation Incorporated
Bethesda, Maryland

Letters to the Editor

Dear Sir:

I read with much interest the letter from Mr. Robert L. Birch in the January 1966 issue of *American Documentation.* I was particularly interested in his general rule No. 5 in the "Bibliographic Suicide-Guide: For Publishers and Authors," included in his letter. The *ASTM Bulletin* was discontinued at the end of the year 1960. The name of the Society was changed from American Society for Testing Materials to American Society for Testing and Materials late in 1961. The change in the Society's name has nothing to do with the unfindability of the *ASTM Bulletin* in view of its discontinuance. In actual fact, the *ASTM Bulletin* which was published eight times a year was replaced by a new monthly publication, MATERIALS RESEARCH & STANDARDS, the first issue of which appears in January 1961.

T. A. Marshall, Jr.
Executive Secretary
American Society for Testing and Materials
Philadelphia, Pennsylvania

Dear Sir:

Mr. Thomas A. Marshall, Executive Secretary of the American Society for Testing and Materials, has very kindly sent me a copy of his letter about the notes on bibliographic suicide technique which appeared in the January 1966 issue of *American Documentation.*

Mr. Marshall is right, of course, in that the 1961 change in the name of his society (from American Society for Testing Materials) could not affect previous efforts to find the *ASTM Bulletin,* which had already ceased publication. His letter brings out a very important point, however, concerning the "labeling" of publications for efficient retrieval from the informational mainstream.

If Mr. Marshall means that the present findability of the very valuable material in the *ASTM Bulletin* remains unaffected, he is only partly right. Each change in the name of a journal or in the name of the sponsoring body silts over one of the strings by which the publication may be traced. In this case, efforts to find still-valuable articles in the *ASTM Bulletin* will be unaffected to the degree that they depend on the title as a tracing. Such a title, beginning with a set of initials, offers more opportunity for confusion in any alphabetical listing than would, for instance, such a title as *Steel Research Journal,* which is easy to remember and offers only one filing location. The *ASTM Bulletin* may be reasonably listed just before such a title as *A to Z in Alphabeting,* or between the *ASPR Handbook* and the ASTME Membership List; all of these, according to some filing traditions, would file before the Aardvark Foundation. Another place to look for the *ASTM Bulletin* would be between *Asters and Asteroids* and, say, *Astounding Science Fiction.* None of these locations, of course, would be affected by a change in the interpretation of the initials or the name of the sponsoring body.

On the other hand, the effort to find an *ASTM Bulletin* article may be through the name of the Society. A searcher looking under the present name of the ASTM would, of course, find no entry for the bulletin, since it ceased publication before the change. If the searcher knows that the Society has changed its name and that the publication ceased before that change, he may seek under the former name. In any case, retrieval of *ASTM Bulletin* material has been affected by the change of name of the Society.

I can sympathize, of course, with the editor who is impatient to get the substantive part of his work done, by getting the information in print and out to the subscribers.

Forty years ago files were smaller and cross references could be made, at least in the larger files, so that a subscriber wishing to put in an order for a hearsay title would be quite likely to be able to identify it and send his order to the right place. The larger the files grow the more necessary it becomes that management study the bibliographic consequences of the wording and layout of titles and the other identification elements connected with a publishing venture.

In the last few months two major journals on materials have been announced. The title of one is *Journal of Materials Science.* The other was announced as *Journal of Materials,* but the illustration of the title page shows it as the *ASTM Journal of Materials.* An earlier publication, still being published, is *Materials Research & Standards;* but, considering the layout of the cover it will probably be referred to as *Materials,* with the Research & Standards being treated as a subtitle. Possibly the title will be read *Materials ASTM,* with the rest as subtitle.

The rich increase in productivity in America made possible by standards such as those prepared so painstakingly by ASTM has made possible a vast growth in the number and diversity of publication efforts. The efforts of directories publishers, and catalogers, and of manual and computer bibliographers, to keep track of the publications and make them findable, will be most successful for those which pose least problems of alternative filing location or memory strain.

The checklist of rocks and shoals on which a new publishing venture may be wrecked (the "Bibliographic Suicide-Guide") which assumed that publishers were parties to a benevolent plot (to prevent searchers from finding the information on which they might have to take action) was designed to permit the launchers of new publications to decide, with better prevision, whether to have their efforts result in new sandbars or whether the new effort will become part of the informational mainstream, findable on demand.

Robert L. Birch
Science Index Group
Falls Church, Virginia

Dear Sir:

I would like to add some remarks about the article of Mr. Harry Baum, A Clearinghouse for Scientific and Technical Meetings: Organizational and Operational Problems, which appeared in *American Documentation,* Vol. 17, No. 1, 28 (1966).

Exactly as the number of journals in all disciplines of science and technology is growing ever more rapidly, so the number of meetings held is increasing, too. It is understandable that oral, direct communications, and personal contacts are excellent for each group of specialists and give more positive results than even the best papers in the journals.

Organizing a clearinghouse for meetings is an excellent idea and I agree with the author that here the cooperation between the organizers of meetings and the clearinghouse is the most important point. The success of the clearinghouse depends on the good will of the organizer. For instance, the efficiency of the clearinghouse could be greatly enhanced if the organizers of a meeting provided the clearinghouse with advance copies of the abstracts of papers to be presented at the meeting. Such abstracts are usually submitted to the organizers at least half a year before the meeting takes place.

I see a second task for the clearinghouse in the publication of proceedings of the meetings. This could be done together with the organizers and specialists chosen by them to evaluate the papers so as to avoid publishing overlapping information. Such a sifting would be of great value to scientists who have to go over the innumerable volumes of proceedings of conferences.

These two functions—quick information about a conference, and publishing activities—should be the goals of the future clearinghouse for scientific and technical meetings.

Alina Gralewska
Head, Library & Technical Inf. Dept.
Soreq Nuclear Research Center

Dear Sir:

I present for your consideration as a note in *American Documentation* a definition that appears in an unpublished paper that I presented at Dillard University (New Orleans, Louisiana) in March of this year. The title of the paper is "Implications of the Great Society for the Natural Sciences."

"Reflective Thinking: Creative information retrieval"

Mahlon C. Rhaney
Dean, College of Arts and Sciences
Florida Agricultural and Mechanical University
Tallahassee, Florida

Dear Sir:

In your double role as editor of *American Documentation* and member of the staff of the leading publisher of citation indexes, you are in a peculiarly effective position to attempt a modest reform in the habits of most scientists and many documentalists.

I refer to the use of initials instead of full forename. I should think this could wreak havoc with citation indexing when one gets to the point that there are 50 different authors all with the name "R. Jones."

I would guess that this practice originates from the association of a few hundred specialists who all know one another in an "invisible college." I also suspect a certain amount of arrogance — "Everyone should know who *I* am." Of all people, documentalists, who sometimes must index millions of citations, should not be guilty of this practice.

R. Jordan
(excuse me, Robert Thayer Jordan)
Council on Library Resources, Inc.
Washington, D.C.

Book Reviews

7/66–IR **INTREX: Report of a Planning Conference on Information Transfer Experiments.** September 3, 1965. Carl F. J. Overhage and R. Joyce Harman, ed. MIT Press, Cambridge, Mass. 276 pp.

The establishment of a four-year program of "information transfer experiments" is M.I.T.'s proposal for finding solutions to the growing pains faced by large research libraries. To set the stage, and formulate a coordinated program, Professor Carl F. J. Overhage and the other INTREX planners organized a five-week planning conference attended by librarians, documentalists, scientists, engineers, publishers, and others. The actual discussions and work sessions were held from August 2 to September 3, 1965; the place was the National Academy of Sciences Summer Studies Center at Woods Hole, Massachusetts. The five-week conference was sponsored by the Independence Foundation of Philadelphia, Pennsylvania.

The report's opening statement describes the objective of the INTREX experiments as follows: ". . . to provide a design for evolution of a large university library into a new information transfer system that could become operational in the decade beginning in 1970." The timetable for the series of experiments, and follow-on installation and implementation, allows four years for experimentation and two or three years for development and construction, so that the target operation time probably centers around 1975. With target times set, and more or less specific investigative goals, it is clear that the M.I.T. effort should not be thought of as a continuing library research effort, but rather as a carefully-limited development project. The care shown by the planners in insuring the realism of the experiments (selection of a real library environment, gathering of a corpus of machine-readable data from actual operating sources) supports the feeling that the experiments will yield practical constructive results.

About half of the INTREX report's 276 pages is devoted to the report text; the other half consists of twenty papers prepared by conference participants on various related subjects. The report text comprises a summary section and seven chapters. The first six chapters develop a picture of the INTREX concept of the future on-line intellectual community, and the last chapter, in seven sections occupying 80 pages, details the proposed INTREX experimental program.

In the first section, discussing the Model System, the contrast is drawn between most traditional library catalogs, which include only the "barest minimum" of information, and the proposed INTREX "augmented catalog," which will not only include additional information such as table of contents, abstract, intellectual level, etc., but also information *about* the catalog itself, such as "records of its own use and use of the documents recorded in it." The catalog would also include data about unpublished works and linkages to other files. The remainder of the discussion of the augmented catalog deals with file organization, search, equipment requirements, and other questions for investigation. Finally, under "text access," the many variations of forms, media, uses, and environmental conditions are suggested for consideration during experiment design.

An all-too-short (5 pages) section entitled "Integration with National Resources" explores the possibility and advisability of making available the resources of a large number of libraries and information centers to individual users. The initial INTREX experiment along this line would tie in with two major specialized information centers — National Library of Medicine and the NASA information system. Other users, both specialized and general, are also contemplated, to add scope and experimental validity.

A section on Fact Retrieval proposes research of more enlarged breadth of vision. In fact, this area of study seems to belong to a generation different from that of the other INTREX experiments. It is proposed to organize the contents of whole groups of basic reference works so as to provide factual responses to questions, rather than references to works containing such responses. By combining the indexing, manipulative, and computational power of the general-purpose computer, far richer and more detailed inquiries would be susceptible to useful treatment. The automated index, automated handbook, and even automated notebook — in which a researcher would record individual and interim notes — are briefly discussed. The implications for the future, following along these lines of research, are only mentioned, but the view from this height begins to make one dizzy, with its suggested implications of availability of all recorded knowledge!

The next section, devoted to initial INTREX facilities, comes back to earth, with the description of the current and soon-to-be-installed computation facilities at M.I.T. (GE 645 and IBM 360/67), and other support, both "hard" and "soft." The recommended initial hardware complement includes a time-shared computer system, and a hierarchy of slower, larger-volume storage devices, such as magnetic-disc, magnetic-chip, and image-microform stores. There will be coaxial cables to the 10 to 30 subscribers, flexible, high-resolution CRT displays and interaction consoles, portable, personal, book-size microform viewers, and remote-inquiry type terminals (typewriters and advanced typewriters).

A series of experiments and speculations are discussed in a section entitled "Related Studies: Extensions and Elaborations." More or less tied to the role of the on-line library in educational processes at the university are: provision of classroom back-up information, self-education, selective dissemination of information, and browsing. Particularly interesting are the suggested experiments to test the validity and utility of "accidental discovery" through conventional and "electronic" browsing. Experiments with use of on-line information to facilitate publication are proposed, including cooperative endeavors by editors, reviewers, and authors while drafts are "in the system." Publication through a microform system, even while in rough manuscript form, is discussed, and situations under which this is advantageous are mentioned. Also briefly mentioned is the notion of "on-demand" publication, whereby on-line text would be reproduced on request, either from microform or digitally-recorded data. Finally, experiments are suggested to investigate methods for reaching decisions on weeding out (or selective retention).

Areas for research and development in support of the INTREX objectives are: consoles, interaction language, content and user needs analysis, and information transfer theory. Research and development obviously deserves extended treatment, but the section on R and D goes into the field only enough to whet the appetite.

Finally, a section on Data Gathering for Evaluation suggests several ways in which the system can be made to keep its own records. They are discussed under the following headings: data on use, economic controls, data on learning, and an annotated user's card.

In the twenty papers (it is understood that these were selected from more than a hundred) comprising the last half of the report, some discuss concepts which found their way into the recommended experimental program. The subjects discussed include network experiments, user studies, graphics, browsing, content-analysis, indexing, education, interaction languages, measurement, modeling. All are thought provoking. Particularly penetrating were Dr. Vannevar Bush's remarks regarding the tremendous impact on society of successful application of analytic machinery in libraries and information systems.

The INTREX report, with its proposed four-year experimental program, details a wide range of investigations showing great imagination and comprehensiveness. It seems to this reviewer that the planners have laid on more than will be possible to attack in four years. Attempts to go too far in some areas may lead to disappointments which will hurt the orderly progress of library technique researches. At the same time, too little was said about relatively mundane matters such as the need for efficient organization of subject-matter terms, search strategies, and indexing techniques.

It is apparent that, even though INTREX experiments will undoubtedly yield much valuable constructive data and insight for all libraries, the M.I.T. locale and obvious science-and-technology slant tend to overemphasize the fulfillment of current science information needs. The inclusion of proposed "fact retrieval" experiments not only underscores this but would tend to add new, possibly overwhelming, dimensions to the library problem, especially affecting storage requirements. Perhaps the tendency in several recent automation proposals to marry the library and the information system has been hasty, although this may become feasible some day ("information system" is used here in the sense of automatically supplying answers to queries rather than references to sources of answers). Simply to enrich a large library's bibliographic apparatus (installing modern techniques, adding tables of contents, abstracts, additional subject indexing, better service and control, etc.) stretches the state of the art. But it is probably not reasonable within the time-frame under discussion to make complete contents of works available electrically, except in specialized situations.

Assembling an all-star cast. Professor Overhage has focused an impressive concentration of technical skill in a dramatic bid to create an image of the future library. To shepherd over 70 distinguished participants and visitors through five weeks of discussion, experimentation, and creative output speaks volumes for his management savvy and diplomatic skill. Furthermore the delivery, within six weeks (the conference ended on September 3 — I received my copy in mid-October) of a finished, well organized report is evidence of a high order of editorial determination. We can all look forward to a stimulating four years of productive experiments in the information sciences.

SAMUEL S. SNYDER
Information Systems Specialist
The Library of Congress

7/66–2R **Alphabetical Subject Indication of Information.** Vol. 3. 1965. John Metcalfe. Rutgers Series on Systems for the Intellectual Organization of Information. Graduate School of Library Service, Rutgers, the State University, New Brunswick, New Jersey. 148 pp.

Mr. Metcalfe, of the University of New South Wales, has had a long and distinguished career as a reference librarian, library science instructor, and library administrator. His specialty is bibliographic organization, and his previous books in the field of indexing, subject cataloging, and subject classification are already known in the library field.

His most recent work is based on his presentation made before a panel at a recent seminar in the Rutgers University Series on Systems for Intellectual Organization of Information, sponsored by the Graduate School of Library Science.

The basic concept of *Alphabetical Subject Indication of Information* is the intuitive system of known names in known order. Familiar examples can be found in the Library of Congress catalog and the Wilson indexes. According to the author, the basic advantage of alphabetical indexing is that it can be applied to an unlimited number of items without the coding difficulties of some other systems.

However, Mr. Metcalfe finds that the system is in an imperfect stage, since little is being done to improve it and understanding of its basic concepts is vague. Thus, the copying of the National Catalog is often done uncritically for reasons of economy and because of the increased inability of the copyists to be critical.

To prove that alphabetical subject indication can serve many purposes, although the system is in need of improvement, the author devotes more than half of his work to tracing the development of this and other subject indication systems and classification orders. A critical analysis of Cutter, Schwartz, Provost, Anderson, and Kaiser is presented. In the field of classification the Standard Catalog, British National Bibliography, Ranganathan's Colon Classification, Dewey's Decimal Classification, Universal Decimal Classification, and the British Technology Index are examined. After the introduction to Alphabetical Subject Indication and the chapter on Historical Background, the book is divided into six additional chapters: III, Input to the System; IV, The Store to be Searched; V, Searching Methods and Output; VI, Discussion of Applications for which the System is Theoretically Suited or Unsuited; VII, Evolution of the Method; and VIII, Seminar Panel Discussion.

Undoubtedly, Mr. Metcalfe's main contribution is his logical defense of the system supported by historical facts and the comparative analysis of other systems. In addition, his chapter on tools for the construction of indexes is well organized.

However, his discussion of chain indexing is at times confusing. Moreover, the author, while aiming at a comprehensive treatment of the subject, did not include recent research in this field. Computer-produced indexes, with the exception of Key Word in Context indexes, are also omitted. Nonetheless, both documentalists and librarians will welcome the appearance of this report as an excellent guide to improvements in alphabetical subject indexing.

GEORGE I. LEWICKY
H. W. Wilson Company

7/66–3R **Principles of Automated Information Retrieval.** 1965. William F. Williams. The Business Press, Elmhurst, Ill. 439 pp.

In *Principles of Automated Information Retrieval,* Mr. William F. Williams sets himself noble and laudable goals. As the introduction states: "This book is intended to eradicate an imaginary and rapidly disappearing boundary line between data processing systems and information retrieval systems. For the executive of either a large or a small organization, it will serve as an introduction to information retrieval systems. . . . For librarians, systems-procedures personnel, systems analysts, and programmers, it will serve as an instruction manual in the design and operation of information retrieval systems. For the executive responsible for business functions which demand improved management and control of information, this book encourages decisions to improve these functions and furnishes adequate authority and background. . . ." Mr. Williams' purpose, then, is to provide THE ONE source book of necessary information for the many varied approaches to mechanized information systems, and to answer the needs of management, librarians, systems and procedures writers, systems analysts, and programmers with one volume. The fact that, in my judgment, he does not succeed in this endeavor does not change my feeling that this is a worthwhile book.

Mr. Williams, who has had considerable information systems experience with a number of industrial organizations, including the pioneering efforts at DuPont, and who is presently Manager of General Electric's Marketing Information Systems, certainly knows the area about which he is writing, and his book gives ample evidence of the fact that his knowledge is based on real, firing-line experience in the design and operation of information systems, and not on vague, specialized, and highly theoretical conclusions. Moreover, he has compiled the information for his work with a great deal of care and effort. The 439 pages are replete with diagrams, flow charts, photographs, and descriptions of techniques and equipment. The 27-page glossary is a combination of definitions from the library, documentation, and data processing fields, and the bibliography and index are both carefully and satisfactorily done.

I suppose the reason for my qualms about the book is that it tries to be too many things for too many people, and it doesn't quite succeed in pulling the rabbit out of the that. In the early chapters Mr. Williams tries to write for the busy executive confronted with the decision of whether or not to automate. Chapter 1, in fact, is optimistically entitled "What Executives Should Know About Initiating New Information Retrieval Systems." I may not be the busy executive at whom Mr. Williams is aiming his message, but I would find the general bits of information contained in the 22 pages of Chapter 1 of little help in "confirming [my] suspicions that the time is ripe to use information retrieval fully." Mr. Williams is an engineer and systems man and his background, of necessity, flavors his writing. If he does not succeed in being all things to all people, I cannot really fault him for this; I doubt that anyone could be. He certainly tries to approach each discipline on its own level of comprehension. The book is full of useful data in tabular and graphical form. For those who presumably don't have other access to it the author even includes a breakdown of the Dewey Decimal system, and a listing of 375 "stop" words for use in Permuted Title Indexing. Nevertheless, as the coverage of the subject proceeds, the mathematical formulas and network charts get thicker, and the going for the non-systems engineer gets rougher.

The author can probably be forgiven his biases in favor of the work of his employer, the General Electric Company. Although other examples are brought in, the book is heavily flavored with GE examples and GE systems. As an example, the permuted title listings are drawn from a 1962 GE manual, and little if any mention is made of the earlier and generally more fundamentally considered work of the late H. P. Luhn. Statements such as "For purely technical information systems, General Electric Co. has made the greatest advance in complete structuring of information" are open to detailed and heated debate, and certainly require greater substantiation than they receive.

Although the purpose is not stated as such, Mr. Williams' book appears to be designed for use as a graduate textbook. Half of the book is given over to chapters concerned with fundamental approaches to documentation, abstracting, indexing, coding, storage, retrieval, and vocabulary control, and all chapters end in a series of questions to test the reader's (or student's) comprehension of the material covered. Terms are carefully defined as introduced, and every attempt is made to proceed in logical sequence. It could almost be assumed that this book was written to serve as a textbook for a course or series of courses on information retrieval to be taught by Mr. Williams and hopefully emulated by others, and this may, in fact, be the case.

Despite its shortcomings, the book is well written, basically well documented, and full of all kinds of information of value to those of us who work in this field. It makes a worthwhile addition to our reference shelves and to those of our technical libraries.

HERBERT S. WHITE
Executive Director
NASA Scientific and Technical Information Facility

7/66-4R International Affairs. Universal Reference System, Political Science, Government and Public Policy Series. Vol. I. 1965. Metron, Inc., 80 East 11th Street, New York. 1205 pp.

No one will argue the fact that lack of bibliographical control coupled with an ever increasing flood of literature has plagued social scientists for decades. Efforts of varying magnitude to cope with the literature access problem have been projected, discussed, and in some cases, applied. Few, if any, of these projects have produced significant results.

The Universal Reference System (URS), as conceived by Alfred de Grazia, is a computerized documentation and information retrieval system which will attempt to provide multifaceted access to the substantive literature of the social and behavioral sciences through selection, storage, and indexing in depth. The services of the URS are to be made available on an individual basis by means of automatic printout and also in published form.

The volume under review is the first to appear of a projected series of ten volumes, each of which will cover some phase of political science, government and public policy. It contains citations and full annotations to 3,030 books, articles, and documents dealing with all aspects of international affairs, drawn from the literature of economics, sociology, anthropology, psychology, and history as well as political science. The nature of the selection process is somewhat vague, but emphasis is upon those works which exhibit strong methodological characteristics. Titles which are purely evaluative, journalistic, non-empirical, or intuitive have been generally excluded. Although the period covered is primarily the twentieth century, with major emphasis upon the post-World War II era, the "classics" of international affairs have been included. Foreign titles make up only 5 percent of the total works cited and it is hoped that the future will see greater expansion in this area. Despite the lack of a well-defined basis of selection the end result is to be commended. The titles included are of high quality and the annotations which accompany each citation are clear and concise, emphasizing scope, methodology, and findings or conclusions reached.

The major key to the URS is its computer indexing system which is based on a general classification of the methodology and techniques of the social sciences devised by Professor de Grazia. Some 183 standard descriptors make up a "topical and methodological index" to which are added 121 "unique descriptors" which are particularly significant to international affairs, e.g., Nationalism, Nuclear Power, Cold War, and individual geographical names. Each document cited has been assigned from ten to twenty descriptors in order to illustrate all of its important facets.

In the first major part of the volume, the "Catalog" (pp. 1–212), all documents cited are randomly listed, giving for each: full bibliographic information, an excellent annotation, and a listing of all descriptors assigned to the title. The second part, "Index of documents" (pp. 213–1197), is an alphabetical descriptor index with each entry comprised of four columns. Each descriptor is listed in the first column with symbols indicating whether the work cited is a book, a long article, or a short article, accompanied by the date of publication. The second column lists three or four other "critical descriptors" — those which indicate the major facets of the work. Column three lists all other descriptors assigned to the document. Column four indicates the serial number of the document as it is listed in the "Catalog." It should be noted that all descriptors, except cross references, listed in the "Index of documents" appear in truncated or abbreviated form, e.g.,

POL/PAR	—Political party
SELF/OBS	—Self observation
BAL/PWR	—Balance of power
REV	—Revolution
SIMUL	—Scientific model
WOR-45	—World wide to 1945

Other features of the volume include: (1) a table of the standard descriptors arranged in their logical classified sequence giving both truncated form and expanded definitions; (2) a frequency table of all descriptors, arranged alphabetically with page references to the "Index of documents"; and (3) an alphabetical author index.

Utility of the URS system in general and of the international affairs volume in particular will of necessity be measured over the passage of time. It represents a substantial departure from the more conventional formats of bibliographical apparatus and necessitates precision and ingenuity on the part of the user in stating his requirements. Mastery of the classification system and the descriptor list is essential to effective use and scope limitations must always be paramount in the user's mind.

Despite the fact that the new format is awkward to use and seems strange to those accustomed to other tools, the producers of the URS must be congratulated on their pioneering efforts to provide social and behavioral scientists with a new and variegated approach to their monumental information problems.

THOMPSON M. LITTLE
Associate Director of Library Services
Hofstra University

Progress in Information Science and Technology

The 29th Annual Meeting of the American Documentation Institute will be held in Santa Monica, California, on October 3 through 7, at the Miramar Hotel.

PROGRAM OUTLINE

TUTORIAL SESSIONS – October 3, 1966

Information Systems Design – R.M. Hayes – UCLA
Information Center Operations – A. Kent – University of Pittsburgh
Usage of Information – S. Herner – Herner and Co.
Evaluation of Hardware and Software – To be announced
Language Data Processing – H.P. Edmundson – SDC
Development of a Theory – D. Hillman – Lehigh University

STUDENT PROGRAM – October 3, 1966

Special Session – Student Papers
Panel Discussion – Student Chapter Activities
Cocktail Hour

J. Harvey – Chairman Student Membership Committee

PROGRESS REVIEW SESSIONS – October 4-7, 1966

Professional Aspects of Information Science and Technology – R.S. Taylor – Lehigh University
Information Needs and Uses – H. Menzel – New York University
Content Analysis, Specification and Control for Document Retrieval Systems – P. Baxendale – IBM
File Organization and Search Techniques – D. Climenson – U.S. Government
Man-Machine Communication – R.M. Davis – Dept. of Defense
Evaluation of Indexing Systems – C.P. Bourne – Programming Services, Incorporated
Automated Language Processing – R.F. Simmons – SDC
New Hardware Developments – M.E. Stevens – National Bureau of Standards
Information System Applications – J. Baruch – Bolt, Beranek and Newman
Library Automation – D.V. Black – University of California, Santa Cruz
Information Centers and Services – G.S. Simpson – Batelle Memorial Institute
National Information Issues and Trends – J. Sherrod – Atomic Energy Commission

SPECIAL FEATURES

Author Forums
Discussion Groups
Prize Papers
Special Libraries Association Session
Placement Service
Exhibits
Information Theater
Chapter Officers Workshop

Special Interest Groups
Proceedings
Award of Merit
Exhibitor Presentations
Tours
Evening in Disneyland
Buffet Luau

If you're located
in the
united states,
canada or
mexico,
we'll schedule
a seminar
for your
organization
on the
science citation ind
and asca

&

perform
on-site searches.
Just ask us
to arrange it.
write dept. 07-4

TUTE FOR SCIENTIFIC INFORMATION
et Philadelphia Pa 19106 USA

hat. In the early chapters Mr. Williams tries to write for the busy executive confronted with the decision of whether or not to automate. Chapter 1, in fact, is optimistically entitled "What Executives Should Know About Initiating New Information Retrieval Systems." I may not be the busy executive at whom Mr. Williams is aiming his message, but I would find the general bits of information contained in the 22 pages of Chapter 1 of little help in "confirming [my] suspicions that the time is ripe to use information retrieval fully." Mr. Williams is an engineer and systems man and his background, of necessity, flavors his writing. If he does not succeed in being all things to all people, I cannot really fault him for this; I doubt that anyone could be. He certainly tries to approach each discipline on its own level of comprehension. The book is full of useful data in tabular and graphical form. For those who presumably don't have other access to it the author even includes a breakdown of the Dewey Decimal system, and a listing of 375 "stop" words for use in Permuted Title Indexing. Nevertheless, as the coverage of the subject proceeds, the mathematical formulas and network charts get thicker, and the going for the non-systems engineer gets rougher.

The author can probably be forgiven his biases in favor of the work of his employer, the General Electric Company. Although other examples are brought in, the book is heavily flavored with GE examples and GE systems. As an example, the permuted title listings are drawn from a 1962 GE manual, and little if any mention is made of the earlier and generally more fundamentally considered work of the late H. P. Luhn. Statements such as "For purely technical information systems, General Electric Co. has made the greatest advance in complete structuring of information" are open to detailed and heated debate, and certainly require greater substantiation than they receive.

Although the purpose is not stated as such, Mr. Williams' book appears to be designed for use as a graduate textbook. Half of the book is given over to chapters concerned with fundamental approaches to documentation, abstracting, indexing, coding, storage, retrieval, and vocabulary control, and all chapters end in a series of questions to test the reader's (or student's) comprehension of the material covered. Terms are carefully defined as introduced, and every attempt is made to proceed in logical sequence. It could almost be assumed that this book was written to serve as a textbook for a course or series of courses on information retrieval to be taught by Mr. Williams and hopefully emulated by others, and this may, in fact, be the case.

Despite its shortcomings, the book is well written, basically well documented, and full of all kinds of information of value to those of us who work in this field. It makes a worthwhile addition to our reference shelves and to those of our technical libraries.

Herbert S. White
Executive Director
NASA Scientific and Technical Information Facility

7/66-4R **International Affairs. Universal Reference System, Political Science, Government and Public Policy Series.** Vol. I. 1965. Metron, Inc., 80 East 11th Street, New York. 1205 pp.

No one will argue the fact that lack of bibliographical control coupled with an ever increasing flood of literature has plagued social scientists for decades. Efforts of varying magnitude to cope with the literature access problem have been projected, discussed, and in some cases, applied. Few, if any, of these projects have produced significant results.

The Universal Reference System (URS), as conceived by Alfred de Grazia, is a computerized documentation and information retrieval system which will attempt to provide multifaceted access to the substantive literature of the social and behavioral sciences through selection, storage, and indexing in depth. The services of the URS are to be made available on an individual basis by means of automatic printout and also in published form.

The volume under review is the first to appear of a projected series of ten volumes, each of which will cover some phase of political science, government and public policy. It contains citations and full annotations to 3,030 books, articles, and documents dealing with all aspects of international affairs, drawn from the literature of economics, sociology, anthropology, psychology, and history as well as political science. The nature of the selection process is somewhat vague, but emphasis is upon those works which exhibit strong methodological characteristics. Titles which are purely evaluative, journalistic, non-empirical, or intuitive have been generally excluded. Although the period covered is primarily the twentieth century, with major emphasis upon the post-World War II era, the "classics" of international affairs have been included. Foreign titles make up only 5 percent of the total works cited and it is hoped that the future will see greater expansion in this area. Despite the lack of a well-defined basis of selection the end result is to be commended. The titles included are of high quality and the annotations which accompany each citation are clear and concise, emphasizing scope, methodology, and findings or conclusions reached.

The major key to the URS is its computer indexing system which is based on a general classification of the methodology and techniques of the social sciences devised by Professor de Grazia. Some 183 standard descriptors make up a "topical and methodological index" to which are added 121 "unique descriptors" which are particularly significant to international affairs, e.g., Nationalism, Nuclear Power, Cold War, and individual geographical names. Each document cited has been assigned from ten to twenty descriptors in order to illustrate all of its important facets.

In the first major part of the volume, the "Catalog" (pp. 1–212), all documents cited are randomly listed, giving for each: full bibliographic information, an excellent annotation, and a listing of all descriptors assigned to the title. The second part, "Index of documents" (pp. 213–1197), is an alphabetical descriptor index with each entry comprised of four columns. Each descriptor is listed in the first column with symbols indicating whether the work cited is a book, a long article, or a short article, accompanied by the date of publication. The second column lists three or four other "critical descriptors" — those which indicate the major facets of the work. Column three lists all other descriptors assigned to the document. Column four indicates the serial number of the document as it is listed in the "Catalog." It should be noted that all descriptors, except cross references, listed in the "Index of documents" appear in truncated or abbreviated form, e.g.,

POL/PAR	—Political party
SELF/OBS	—Self observation
BAL/PWR	—Balance of power
REV	—Revolution
SIMUL	—Scientific model
WOR-45	—World wide to 1945

Other features of the volume include: (1) a table of the standard descriptors arranged in their logical classified sequence giving both truncated form and expanded definitions; (2) a frequency table of all descriptors, arranged alphabetically with page references to the "Index of documents"; and (3) an alphabetical author index.

Utility of the URS system in general and of the international affairs volume in particular will of necessity be measured over the passage of time. It represents a substantial departure from the more conventional formats of bibliographical apparatus and necessitates precision and ingenuity on the part of the user in stating his requirements. Mastery of the classification system and the descriptor list is essential to effective use and scope limitations must always be paramount in the user's mind.

Despite the fact that the new format is awkward to use and seems strange to those accustomed to other tools, the producers of the URS must be congratulated on their pioneering efforts to provide social and behavioral scientists with a new and variegated approach to their monumental information problems.

Thompson M. Little
Associate Director of Library Services
Hofstra University

Progress in Information Science and Technology

The 29th Annual Meeting of the American Documentation Institute will be held in Santa Monica, California, on October 3 through 7, at the Miramar Hotel.

PROGRAM OUTLINE

TUTORIAL SESSIONS – October 3, 1966

Information Systems Design – R.M. Hayes – UCLA
Information Center Operations – A. Kent – University of Pittsburgh
Usage of Information – S. Herner – Herner and Co.
Evaluation of Hardware and Software – To be announced
Language Data Processing – H.P. Edmundson – SDC
Development of a Theory – D. Hillman – Lehigh University

STUDENT PROGRAM – October 3, 1966

Special Session – Student Papers
Panel Discussion – Student Chapter Activities
Cocktail Hour

J. Harvey – Chairman Student Membership Committee

PROGRESS REVIEW SESSIONS – October 4-7, 1966

Professional Aspects of Information Science and Technology – R.S. Taylor – Lehigh University
Information Needs and Uses – H. Menzel – New York University
Content Analysis, Specification and Control for Document Retrieval Systems – P. Baxendale – IBM
File Organization and Search Techniques – D. Climenson – U.S. Government
Man-Machine Communication – R.M. Davis – Dept. of Defense
Evaluation of Indexing Systems – C.P. Bourne – Programming Services, Incorporated
Automated Language Processing – R.F. Simmons – SDC
New Hardware Developments – M.E. Stevens – National Bureau of Standards
Information System Applications – J. Baruch – Bolt, Beranek and Newman
Library Automation – D.V. Black – University of California, Santa Cruz
Information Centers and Services – G.S. Simpson – Batelle Memorial Institute
National Information Issues and Trends – J. Sherrod – Atomic Energy Commission

SPECIAL FEATURES

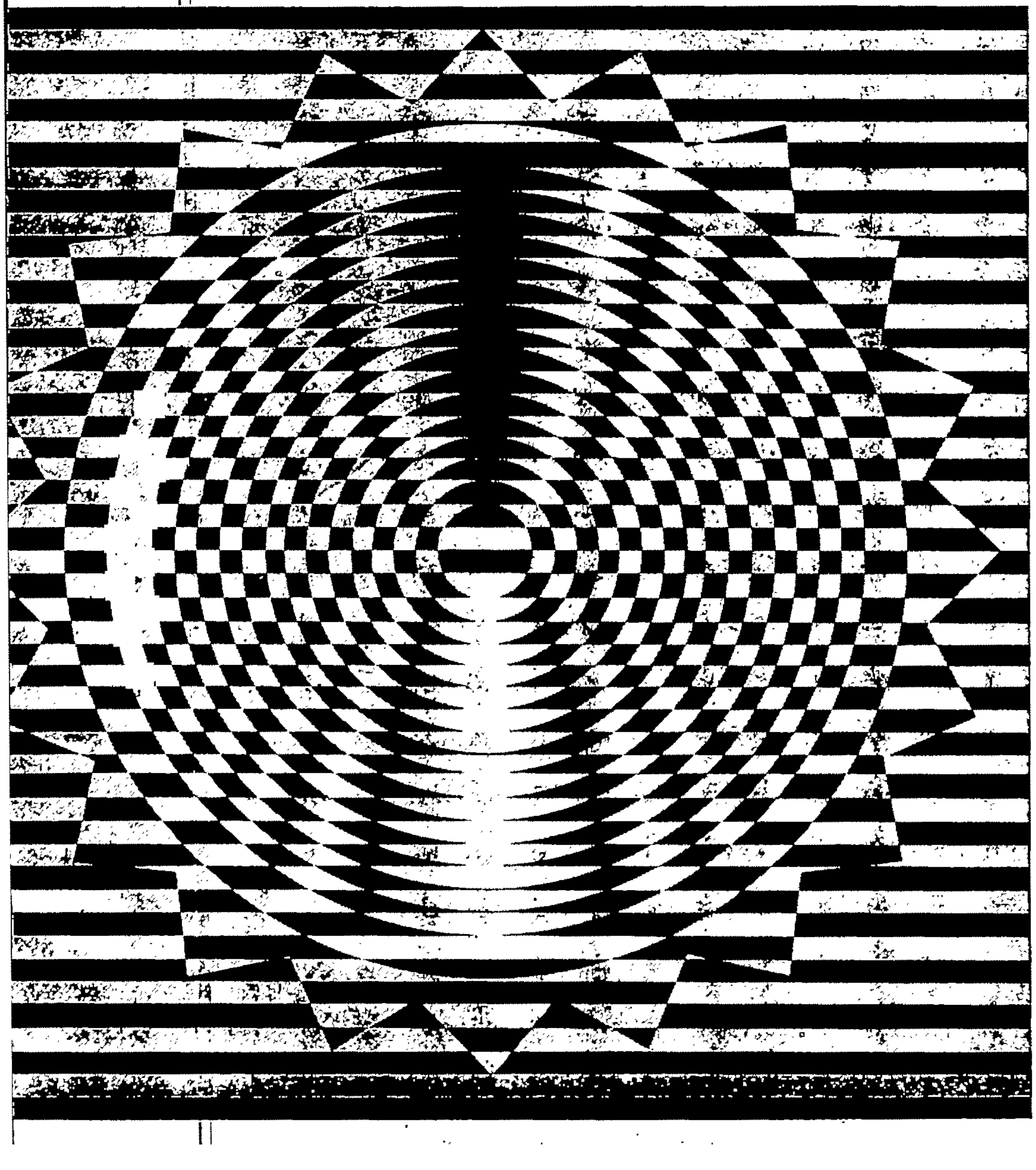
Author Forums
Discussion Groups
Prize Papers
Special Libraries Association Session
Placement Service
Exhibits
Information Theater
Chapter Officers Workshop

Special Interest Groups
Proceedings
Award of Merit
Exhibitor Presentations
Tours
Evening in Disneyland
Buffet Luau

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

OCTOBER
1966
Vol. 17, No. 4



29TH
ANNUAL
MEETING
AMERICAN
DOCUMENTATION
INSTITUTE
OCTOBER 3-7
MIRAMAR HOTEL
SANTA MONICA
CALIFORNIA

AMERICAN DOCUMENTATION

INSTRUCTIONS TO AUTHORS

American Documentation is a publication of the American Documentation Institute. It is a scholarly journal in the various fields in documentation and serves as a forum for discussion and experimentation. Papers already published or in press elsewhere are not acceptable. For each proposed contribution, one original and two copies (in English only) should be mailed to Mr. Arthur W. Elias, Editor, *American Documentation,* Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106. The manuscript should be mailed *flat* in a suitable-sized envelope. Graphic materials should be submitted with suitable cardboard backing.

Types of Manuscripts: Three types of contributions are considered for publication: full-length articles, brief communications of 1,000 words or less, and letters to the editor. Letters and brief communications can generally be published sooner than full-length manuscripts. Books, monographs, and reports are accepted for critical review. Two copies should be addressed to the Review Editor, Dr. T. Hines, 54 North Drive, East Brunswick, New Jersey.

Processing: Acknowledgment will be made of receipt of all manuscripts. *American Documentation* employs a reviewing procedure in which all mansucripts are sent to two referees for comment. When both referees have replied, copies of their comments are sent to authors with the Editor's decision as to acceptability. The refereeing procedure requires about 30 days. Authors receive galley proofs with a five-day allowance for corrections. Standard proofreading marks should be employed. Reprint order forms are forwarded with galleys.

Format: All contributions should be typewritten on white bond paper on one side only, leaving about 1.25 inches (or 3 cm) of space around all margins of standard, letter-size (8.5 × 11 inch) paper. Double spacing must be used throughout, including the title page, tables, legends, and references. The first page of the manuscript should carry both the first and last names of all authors, the institutions or organizations with which the authors are affiliated, and notation as to which author should receive the galleys for proofreading. All succeeding pages should carry the last name of the first author in the upper right-hand corner (0.5 inch from the top) and the number of the page.

Style: In general, style should follow the forms given in the Style Manual for Biological Journals (SMBJ), published for the Conference of Biological Editors by the American Institute of Biological Sciences (1964).

Title: The title should be as brief, specific, and descriptive as possible. Vague and unrevealing titles may delay publication.

Abstract: An informative abstract of 200 words or less must be included, typed with double spacing on a separate sheet. This abstract should present the scope of the work, methods, results, and conclusions.

Acknowledgments: Financial support may be listed as a footnote to the title. Credit for materials and technical assistance or advice may be cited in a section headed "Acknowledgments," which should appear at the end of the text. General use of footnotes in the text should be avoided.

Graphic materials: *American Documentation* requires finished artwork. Follow the style in current issues for layout and type faces in tables and figures. A table or figure should be constructed so as to be completely intelligible without further reference to the text. Lengthy tabulations of essentially similar data should be avoided.

Figures should be lettered in black India ink. Charts drawn in India ink should be so executed throughout, with no typewritten material included. Letters and numbers appearing in figures should be distinct and large enough so that no character will be less than 2 mm high after reduction. A line 0.4 mm wide reproduces satisfactorily when reduced by one-half. Graphs, charts, and photographs should be given consecutive figure numbers as they will appear in the text; however, figure numbers and legends should not appear as part of the figure, but should be typed double spaced on a separate sheet of paper. Each figure should be marked *lightly* on the back with the figure number, author's name, complete address, and shortened title of the paper.

For figures, the originals with two clearly legible reproductions (to be sent to referees) should accompany the manuscript. In the case of photographs, three glossy prints are required, preferably 8 × 10 inches.

Organization: In general, papers should state the background and purpose of the study, followed by details of methods, materials, procedures, and equipment. Findings, discussion, and conclusions should appear in that order. Appendixes may be employed where appropriate for extensive lists, statistics, and other supporting data.

Bibliography: Accuracy and adequacy of the references are the responsibility of the author. Therefore, literature cited should be checked carefully with the original publications. References to personal letters, abstracts of verbal reports, and other unedited material may be included. If an as-yet-unpublished paper would be helpful in the evaluation of a manuscript, it is advisable to make a copy of it available to the Editor. When a manuscript is one of a series of papers, the preceding member of the series should be included in literature cited.

Citation Format:

Order: Literature cited should be sequentially numbered as cited.

Authors: Give all authors with arrangement as follows:
Elias, A. W., B. H. Weil, and I. D. Welt

Titles: Give full titles of articles in English, indicating language of original as: (In Ger.)

Journals: Journal titles should be given in full.

Monograph and Serial Data: Should be presented in order as follows: Volume, issue number, pagination, and year. The issue number should be given in parentheses if journal pagination is not continuous from issue to issue. Pagination should be inclusive. Year of publication should be given in parentheses. An example is given below:

Bishop, D., A. L. Milner, and F. W. Roper, Publication Patterns of Scientific Serials, American Documentation, 16 (No. 2): 113–21 (1965).

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), three copies in the contributing membership fee ($100.00 per year). and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts, Science Citation Index, Chemical Abstracts,* and *Documentation Abstracts.*

American Documentation is entered for second class mailing at Baltimore, Maryland.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 17, No. 4 OCTOBER 1966

Published quarterly by American Documentation Institute and printed by Port City Press, Inc., 1323 Greenwood Road, Baltimore, Maryland 21208. Second class postage paid at Baltimore, Maryland.

Editorial

WHAT IS THE STATE OF THE ART?

Recent conferences and institutes directed by your Editor and Associate Editor have offered more-than-ample support for a feeling that I have had for a number of years. People have practical problems and expect help in solving them. Sophisticated theoretical constructs of nonexistent, ideal situations do not help those engaged in the day-to-day activities of special libraries and information centers. Neither do the conventional librarianship courses offered at most of our library schools.

How do we help the neophyte to learn what has gone on before? The critical, authoritative "state-of-the-art" review is one possible solution; the book review is another. These are sorely needed, particularly in the area that I have characterized as "science information" (Editorial, *Am. Doc.*, Oct. 1964). After all, most of our members and readers are actively engaged in the development and maintenance of on-going information services, and they want to know how others have solved problems similar to those that they encounter day-by-day.

It is very encouraging to note that within recent months articles and papers directed to the solutions of some of these problems have indeed appeared with increasing frequency. Our sister publication, *Special Libraries,* for example, is to be congratulated upon its editorial policy of encouraging such contributions and of publishing them.

American Documentation, too, has been fortunate in receiving an increasing number of excellent, practical state-of-the-art papers. Excellent examples of these may be found in the masterful syntheses by Marguerite Fischer and John Markus in the October 1965 issues of our journal. My students are very grateful. Our thanks, too, to Ted Hines and his Columbia University students for their exceptionally good book reviews. With a plethora of proceedings publications (e.g., Drexel TICA and American University's Technology of Management Series), critical book reviews may often be the "infanticidal agents" for brain children refractory to editorial "birth control" methods.

Isaac D. Welt, Ph.D.
Associate Editor, *American Documentation*
Deputy Director for Scientific
and Technical Information Systems
Center for Technology and Administration
The American University
Washington, D.C. 20006

Comprehensive Dissemination of Current Literature*

A résumé of the Ames Laboratory Selective Dissemination of Information (SDI) System is presented and its potential for future generation computers is discussed. The system is compared with other operational SDI systems with particular emphasis on the design differences. The Ames Laboratory system's adaptability to different input tape compositions and subject coverage is shown through a study of the results of 26 production runs made from two distinct document sources. A detailed analysis of the Ames Laboratory SDI System is made for a 40-run period in 1965, including a discussion of shortcomings of the system and suggested solutions to eliminate certain areas of "noise."

C. R. SAGE

Institute for Atomic Research
Iowa State University
Ames, Iowa

• I. Introduction

Systems, such as the Ames Laboratory Selective Dissemination of Information System (SDI), are commonplace at many installations and they must be evalulated accurately and the results disseminated so that the design criteria of these and future systems are continually improved. The results described are unique because they convey a segment of the scientific community's reaction to an operational computerized information system, while the designers of the programmed system are only bystanders measuring and evaluating these reactions. The system is designed with a minimum of user or document restrictions and adapts to individual users and source documents depending upon user participation.

Section II contains a brief review of the design of the Ames Laboratory SDI System (1). The differences in design from other current awareness systems are emphasized. A detailed explanation is given concerning profile design and the "decision-making" algorithms coupled with comments concerning the choice of these particular methodologies. Section III is a comparative study of the system in operation using two distinct types of documents and 21 profiles representing 18 individual scientists. The main purpose of this study was to determine the system's ability to adapt to different document sets through mathematical control. The 21 profiles chosen for this study were a segment of 35 profiles which had been employed in SDI since its inception, but had not undergone any major manual revisions.

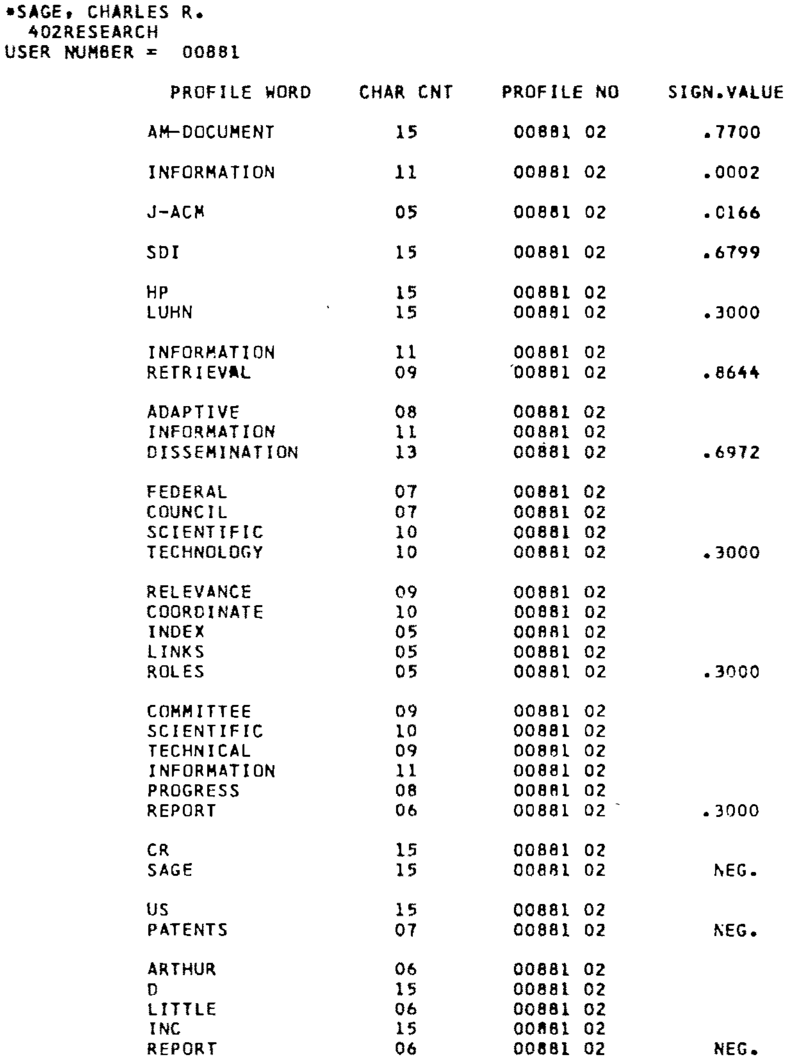
Section IV contains the results of the system while in operation during 1965. The document input used in 1965 actually covered a six-month calendar period. The users and profiles are grouped loosely into the following seven categories: metallurgy, chemistry, experimental physics, theoretical physics, reactor, engineering, and mathematics and computers. The only reason for categorizing profile interests was because the display of the graphic results would be too bulky on an individual profile basis. We have made no attempt to screen out poor profiles, disinterested users, or poor discipline coverage. Many of the profiles shown had been extensively revised and expanded during this period of operation. The number of document entries is large. One source document purchased from the *Institute for Scientific Information*, Philadelphia, Pa., has a very broad coverage of more than 1,060 scientific journals as indexed in *The Science Citation Index* 1965 and all U. S. Patents. The second source document (2), *Nuclear Science Abstracts*, has a narrower scope, having undergone a human selection process. However, it covers the international literature, including numerous governmental reports on nuclear science and technology, and is perhaps the most comprehensive abstracting service in the nuclear science field.

* Work was performed in the Ames Laboratory of the U. S. Atomic Energy Commission. Contribution No. 1880.

• II. Ames Laboratory SDI Computer System

Interest profiles consist of single words or groups of two to six words, each with a significance value ranging betwen 0 and 1 to four significant figures. Single words and groups of words are statistically independent in the matching process; words within each group imply that all words must be in a given document or the significance value of that word group is not considered. Truncation and extension capabilities of words and groups of words are available to the user. Words and groups of words can be negated. Negative terms are not weighted; they are treated as total negation. A profile is limited to 10,000 words or word groups; a user is limited to 99 profiles.

It has been observed from 83 profiles now employed in SDI that the word composition of profiles varies greatly. The average size of a profile is comprised of 79.8 words or word groups. The range in size of profiles varies from 2 to 1,241 words or groups of words per profile. Users are not restricted to follow a thesaurus, and they have the capability to include in their profiles foreign language terms, journal sources, authors, and any term or combination of terms they feel is pertinent to their interests. We encourage large profiles; however, we have no control over this situation. The majority of participants in SDI have started with comparatively small profiles and some feel these smaller profiles fit their needs. However, the majority of users continue to expand their profiles from week to week. Each individual is responsible for the performance of his profile. He benefits from the system by the amount of effort he expends to fully utilize the potential offered by the system. The system is designed with enough flexibility to accommodate varying needs of users without creating burdensome maintenance problems. Figure 1 is a condensed example of a typical profile.



```
*SAGE, CHARLES R.
  402RESEARCH
USER NUMBER =  00881

            PROFILE WORD      CHAR CNT    PROFILE NO      SIGN.VALUE

          AM-DOCUMENT            15        00881 02          .7700

          INFORMATION            11        00881 02          .0002

          J-ACM                  05        00881 02          .0166

          SDI                    15        00881 02          .6799

          HP                     15        00881 02
          LUHN                   15        00881 02          .3000

          INFORMATION            11        00881 02
          RETRIEVAL              09        00881 02          .8644

          ADAPTIVE               08        00881 02
          INFORMATION            11        00881 02
          DISSEMINATION          13        00881 02          .6972

          FEDERAL                07        00881 02
          COUNCIL                07        00881 02
          SCIENTIFIC             10        00881 02
          TECHNOLOGY             10        00881 02          .3000

          RELEVANCE              09        00881 02
          COORDINATE             10        00881 02
          INDEX                  05        00881 02
          LINKS                  05        00881 02
          ROLES                  05        00881 02          .3000

          COMMITTEE              09        00881 02
          SCIENTIFIC             10        00881 02
          TECHNICAL              09        00881 02
          INFORMATION            11        00881 02
          PROGRESS               08        00881 02
          REPORT                 06        00881 02          .3000

          CR                     15        00881 02
          SAGE                   15        00881 02           NEG.

          US                     15        00881 02
          PATENTS                07        00881 02           NEG.

          ARTHUR                 06        00881 02
          D                      15        00881 02
          LITTLE                 06        00881 02
          INC                    15        00881 02
          REPORT                 06        00881 02           NEG.
```

FIG. 1. Ames Lab SDI user profile listing

In the matching process of document words and profile words, only the number of characters specified in the profile word records is matched. All words have a maximum of 15 characters. All of the words in a document entry are used in the matching process with the exception of a 210-word dictionary file of articles, prepositions, etc. In the original systems design of our SDI system we could not afford to create or transcribe our own machine-readable documents. Consequently, we programmed the "front-end" of the system to accept any type of machine-readable input. The "front-end" program is capable of scanning and creating individual word records from six various types of document formats: *Science Citation Index, Chemical Titles, Sandia Corporation Publications Accession Lists, Nuclear Science Abstracts, IBM KWIC Index,* and *Ames Laboratory Publication Master File.* This program was written in modular form to compensate for the addition or deletion of formats.

We are confronted with an upper limit due to computer memory size. This "open-ended" design feature has allowed the system to operate on a production basis in a relatively short period of time since we have not saddled our installation with preparing document input. We are in the delightful position of shopping for various machine-readable source documents of input which would be pertinent to our scientists' research interest areas. Government, industry, etc., are now starting to make available various types of tape information files, which should broaden the scope of our current awareness computer service. Our present input consists of 6,500–7,500 document entries per week and is running approximately three and a half hours per week on an IBM 7074/1401 12 tape, 20K computer. Expansion of literature coverage would not be an expensive item taking into account the broad scientific literature coverage our scientists would have at their disposal.

The "decision making" for selection has been designed to compensate for various source documents of different composition (titles and authors only; titles, authors, sources, and abstracts, etc.). This compensation has been achieved through varying threshold levels for uniquely composed document sets at the time the summation is made of significance values to determine if document entries should be transformed into notifications. The sum of significance values of words and word groups within a given document entry unique to a particular user is obtained by use of the formula for the probability of the union of two events. If we let T_K be the total probability that the user will want a particular document entry, the function below progressively calculates T_K.

$$O < P_K < 1 \text{ (Probability of Word or Word Group Significance Value)}$$

$$T_K = T_{K-1} + P_K - P_K T_{K-1} \text{ (Summation Function)}$$

The threshold value mentioned above is compared with the sum of the significance values. If it is greater than the sum, a notification is not generated for a user.

After receiving the notification, the user specifies his interest in a particular document by pushing out a "Port-a-Punch®" option on a response card, which is attached to the notification containing the original document information. The options available to the user are: "OF INTEREST, DOC. REQUESTED," "OF INTEREST, DOC. NOT REQUESTED," "IMPARTIAL, DON'T ADJUST PROFILE," "OF NO INTEREST," "THE USER ABOVE IS ABSENT." The profile words and word groups which interacted with equivalent document entry words and word groups are increased, decreased, or not adjusted, depending on the types of response options punched and their frequency distribution within the total document set of a particular run.

There are two distinct feedback mechanisms. A normal feedback function (3) involves increasing, decreasing, or not adjusting term significance values as demanded by user responses. If response cards are not returned within a prescribed time limit (four weeks) from the date of distribution, they are treated as negative or "OF NO INTEREST." We have found this to be very valuable because distinterested users (and every similar service of this sort will have them) are gradually starved for notifications. This "cut-off" has also aided us in compiling statistics on all notifications generated by the system.

An abnormal feedback function (3) is also used in order to stabilize the system during transient fluctuations caused by abnormal document distribution to allow for interest changes and to allow for renewed or reformed interest changes. All profile words and word groups which matched equivalent document words and word groups, but for which a document was not selected for a notification, are affected by the abnormal feedback function. The abnormal or slow increment function is based on the supposition that profile words occurring infrequently in document entries of a given dissemination run should be reconsidered by the user, since the projected volume of notifications containing these profile words will be relatively small. Therefore, their respective significance values are incremented similarly, but to a lesser degree (the normal increment and decrement are not derived by a linear function), to a positive response word or word group significance value. Conversely, frequently appearing profile words with low significance values as calculated from past responses are considered general terms and a proportionately small increment or no increment is desired. This abnormal feedback function is designed to limit the increment range within 0 to half that of a normal increment; the value in the range determined by the inverse ratio of the frequency of the word or word group appearing in the document entries of each dissemination run.

It is clear that each word or word group has a unique weight with regard to each individual user and the system has the natural ability to automatically analyze and readjust this weight under complete jurisdiction of the user by his past document selections. With little effort the user acquires a rather sophisticated linguistic analysis tailored completely to the individual rather than to scientific discipline. We have allowed users the option of assigning constant weights excluding the feedback mechanisms, but the few users who attempted this scheme were flooded with indiscriminate information and have converted to the feedback scheme. Actually, a user, who attempts to manually assign weights to profile words and word groups for selection of pertinent current literature from the weekly volume we are processing, will find his task most time consuming and impossible for utilization purposes. A user is prepared to weight terms in relation to his own field of endeavor; however, a problem of selectivity arises with broad-coverage source documents and the user's lack of knowledge of these same terms and the role they play in other scientific fields. Feedback automatically compensates to the great degree of allowing both the user and the system the capabilities of handling many diverse interest profiles efficiently without having to filter the document input into disciplines beforehand and running specialized inhibited source documents autonomously. We feel many SDI systems in operation are not basically selective systems because document entries are preselected before computer dissemination and selection.

We have found that to generate any participation enthusiasm one must be prepared to offer a very comprehensive literature coverage to a prospective user. If a current awareness service cannot honestly supplement the user's present literature capabilities, the task of "selling" the system for service will be very difficult. During a six-month period of production the number of users has increased from 1/3 to 3/4 the total number of users eligible to participate in SDI (restricted to senior scientists of the Ames Laboratory). Based on two "drop-outs" and response cut-offs, 84% of the present users are actively participating in SDI. One of the "drop-outs" was in a very specialized area of research which SDI was not covering with its document input. Participation in SDI is strictly voluntary.

In the program design of the Ames Laboratory SDI system (14 production programs) we attempted to make the system "open-ended." There is no theoretical limit to the actual size of a document, and the upper limit of words and word groups for profiles indicates no hindrance to the user. There are practical limitations to the system which are dictated by computer configurations. Our SDI system is dependent on the sorting capabilities of the computer; however, this inhibitor is not reflected theoretically to volumes but to efficiency of computer running time. The important point is that this system with its present basic systems design has the potential of being universal enough for future generation computers. The systems adaptability to various sources of input certainly generates enthusiasm over the advent of effective optical scanners for application beyond projected routine storage applications. The systems adaptability to individuals rather than to disciplines or specific research areas lends itself to the feasibility of centralized nonspecialized current awareness centers. On-line audio response mechanisms should replace punching response holes; visual display consoles will replace printed notifications; retrospective search capabilities will retrieve reference and cited articles; transformation of diagrams, formulae, photographs, etc., into machine-readable form will expand profile capabilities; and any number of exotic improvements can easily be incorporated into this basic system.†

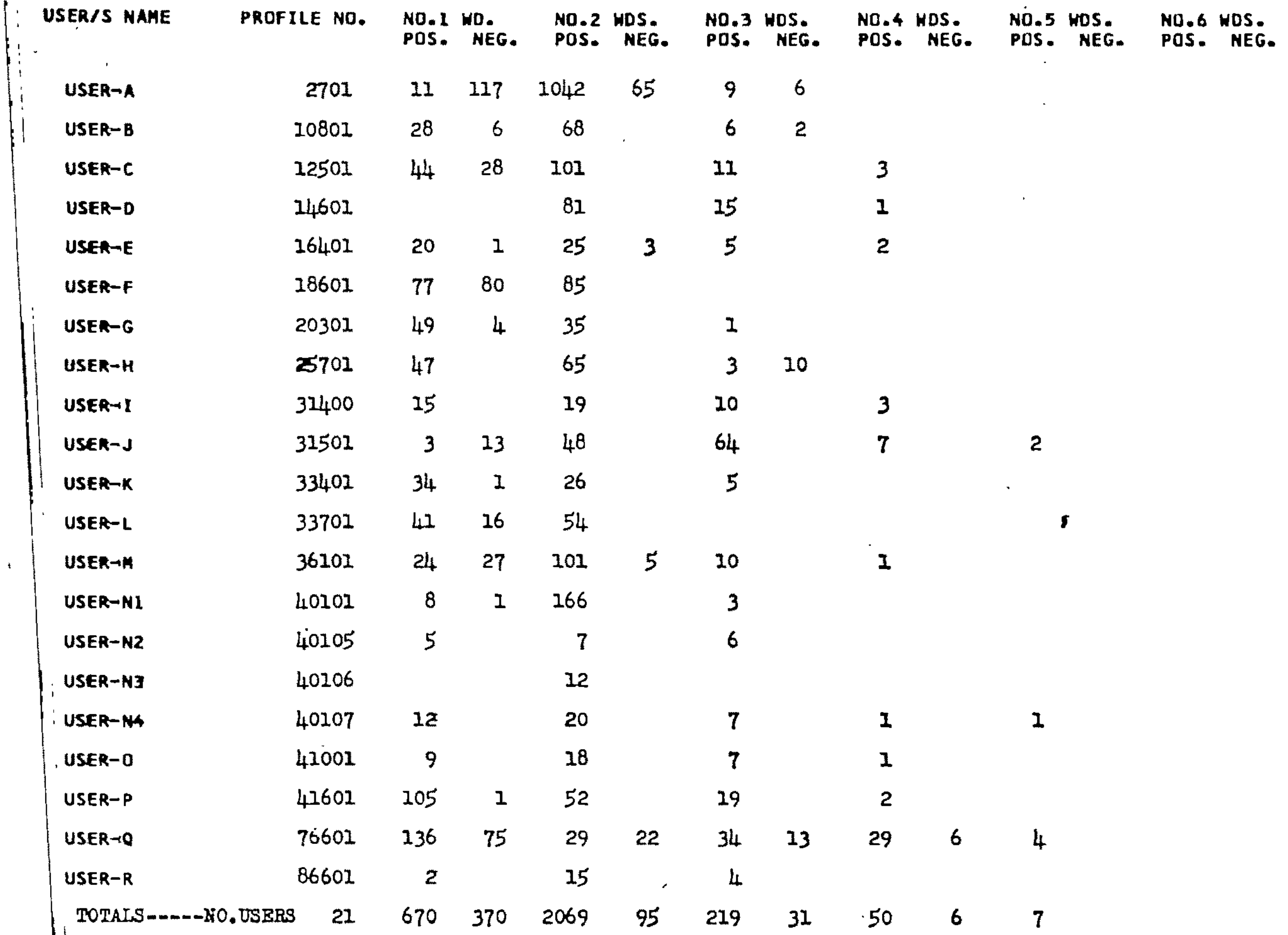
• III. Statistical Measurement of Document Adaptability

The purpose of this study was to measure the relative effectiveness of the Ames Laboratory SDI computer system to adapt to differently composed source documents by variance of threshold values. Twenty-one profiles which had not undergone major manual revisions were picked for this study employing the statistics of 26 production runs. Profiles were chosen strictly on the above criteria; they were not filtered out because of poor design or because they were functioning incorrectly. All 21 profiles had been run in a pilot study prior to production. The significance values of the words and word groups had been partially adjusted as a result of feedback from the pilot study which involved 9,428 document entries. Structuring and word composition of these profiles are quite diversified as shown in Fig. 2.

Corresponding to Fig. 2 are brief résumés of the areas of interest of the individual scientists who compiled the twenty-one profiles. The differentiation of research interests is quite evident and it is important to remember the system was performing a service during these 26 runs and the statistics compiled are the candid responses of these 18 scientists.

The source documents scanned for this study were 13 semi-monthly issues of *Nuclear Science Abstracts* and 13 weekly issues of *Science Citation Index*. *Nuclear Science Abstracts* tape records contained authors, corporate author, title, source, and keywords of Volume 19, Numbers 1, 4, 8–17, 1965. *Science Citation Index* tape records contained authors, titles, and sources of weeks 33–45, 1965. There was a definite overlap in journal article coverage; *Nuclear Science Abstracts* differed by containing USAEC research and development reports; *Science Citation Index* differed by containing all U. S. patents. There were certain prejudices held by the scientists in relation to these source documents. In a recent survey (4) of all our users we found that 43% would prefer to be dropped from *Nuclear Science Abstracts* coverage, and 2% would prefer to be dropped from the *Science Citation Index* coverage. We could not compensate for these discrepancies.

† To be subjected to detailed experimental studies.



```
USER/S NAME    PROFILE NO.   NO.1 WD.    NO.2 WDS.   NO.3 WDS.   NO.4 WDS.   NO.5 WDS.   NO.6 WDS.
                            POS.  NEG.  POS.  NEG.  POS.  NEG.  POS.  NEG.  POS.  NEG.  POS.  NEG.

  USER-A              2701    11   117  1042    65     9     6
  USER-B             10801    28     6    68           6     2
  USER-C             12501    44    28   101          11           3
  USER-D             14601                81          15           1
  USER-E             16401    20     1    25     3     5           2
  USER-F             18601    77    80    85
  USER-G             20301    49     4    35           1
  USER-H             25701    47          65           3    10
  USER-I             31400    15          19          10           3
  USER-J             31501     3    13    48          64           7           2
  USER-K             33401    34     1    26           5
  USER-L             33701    41    16    54
  USER-M             36101    24    27   101     5    10           1
  USER-N1            40101     8     1   166           3
  USER-N2            40105     5           7           6
  USER-N3            40106                12
  USER-N4            40107    12          20           7           1           1
  USER-O             41001     9          18           7           1
  USER-P             41601   105     1    52          19           2
  USER-Q             76601   136    75    29    22    34    13    29     6     4
  USER-R             86601     2          15           4
   TOTALS-----NO.USERS  21   670   370  2069    95   219    31    50     6     7
```

FIG. 2. Analyses of the 21 profiles for computer study

Name	Title	Area of interest
USER-A	Chemist	Mass spectroscopy applied to inorganic and analytical chemistry, corrosion chemistry, high vacuum techniques, special gas handling problems, precision isotope abundances, absolute isotope abundances, isotope geochemistry
USER-B	Chem. Eng.	Thermodynamics of liquid metal systems, kinetics of metal halide reactions, preparation of high purity metals
USER-C	Metallurgist	Mechanical metallurgy
USER-D	Chemist	Physical and inorganic chemistry
USER-E	Ceramic Eng.	High temperature refractory ceramics, high temperature systems and reactions
USER-F	Metallurgist	Physical and mechanical metallurgy, oxidation
USER-G	Metallurgist	Physical and chemical metallurgy
USER-H	Chemist	Analytical chemistry
USER-I	Chem. Eng.	Solvent extraction chemistry
USER-J	Chemist Metallurgist Physicist	Physical chemistry, solid state physics, physical metallurgy
USER-K	Physicist	Theoretical physics and nuclear physics
USER-L	Physicist	Solid state theory
USER-M	Physicist	Nuclear physics
USER-N	Chemist	X-ray and neutron diffraction
USER-O	Physicist	Nuclear physics, accelerator design
USER-P	Physicist	Experimental solid state physics
USER-Q	Metallurgist	Rare-earth metallurgy and solid state physics, alloy theory, high pressure physics
USER-R	Chemist	Surface chemistry, electrochemistry and adsorption, thermodynamics, statistical mechanics

Figure 3 is the graphic result of the *Nuclear Science Abstracts* runs. The top of Fig. 3 is a bar graph indicating the number of neutral (Impartial), negative (No Interest), and positive (Of Interest) responses from top to bottom, respectively, with the numeric quantity of positive responses printed for each run in the corresponding bar. A total of 25,378 document entries made up these 13 runs. Each document entry averaged 43.1 words and each run averaged 1925.15 document entries. The average numbers of responses per run were 260.54 positive, 384.46 negative, and 218.54 neutral for a total of 863.54. The bottom line graph of Fig. 3 illustrates the percentage of positive responses based on the total number of positive and negative responses for each run. We did not take into account, in determining these percentages, the neutral responses because we do not know how to classify them. The line across the line graph is the average percentage of positive responses based on the total number of positive and negative responses for each of the 13 runs. All percentages are relative to notifications disseminated and do not include all potential relevant documents. The average relative percentage of positive responses was 40.39%.

Figure 4 is the graphic results of *Science Citation Index* runs. A total of 69,348 document entries made up these 13 runs. Each document entry averaged 11.2 words and each run averaged 6103.7 document entries. The average numbers of responses per run were 145.85 positive, 209.00 negative, and 109.59 neutral for a total of 464.44. The average relative percentage of positive responses based on the total number of positive and negative responses was 41.10%. Figure 5 (a and b) gives the numeric averages of the total neutral, negative, and positive responses for each user of the two different source documents used. Graphs are available of each individual's results over the 26-run period (5).

A threshold value of .5 was used for *Nuclear Science Abstracts* and a threshold of .3 was used for *Science Citation Index*. In the original design of the increment-decrement feedback function we believed that profile words and word groups with significance values near the .5 threshold level would receive the largest increments and decrements on the assumption that a .5 threshold would be used for all documents. We found through experimentation prior to this study that if the threshold were held constant, significance values of profile words and word groups could not adapt uniformly to document base changes and produce optimum results. The signifi-

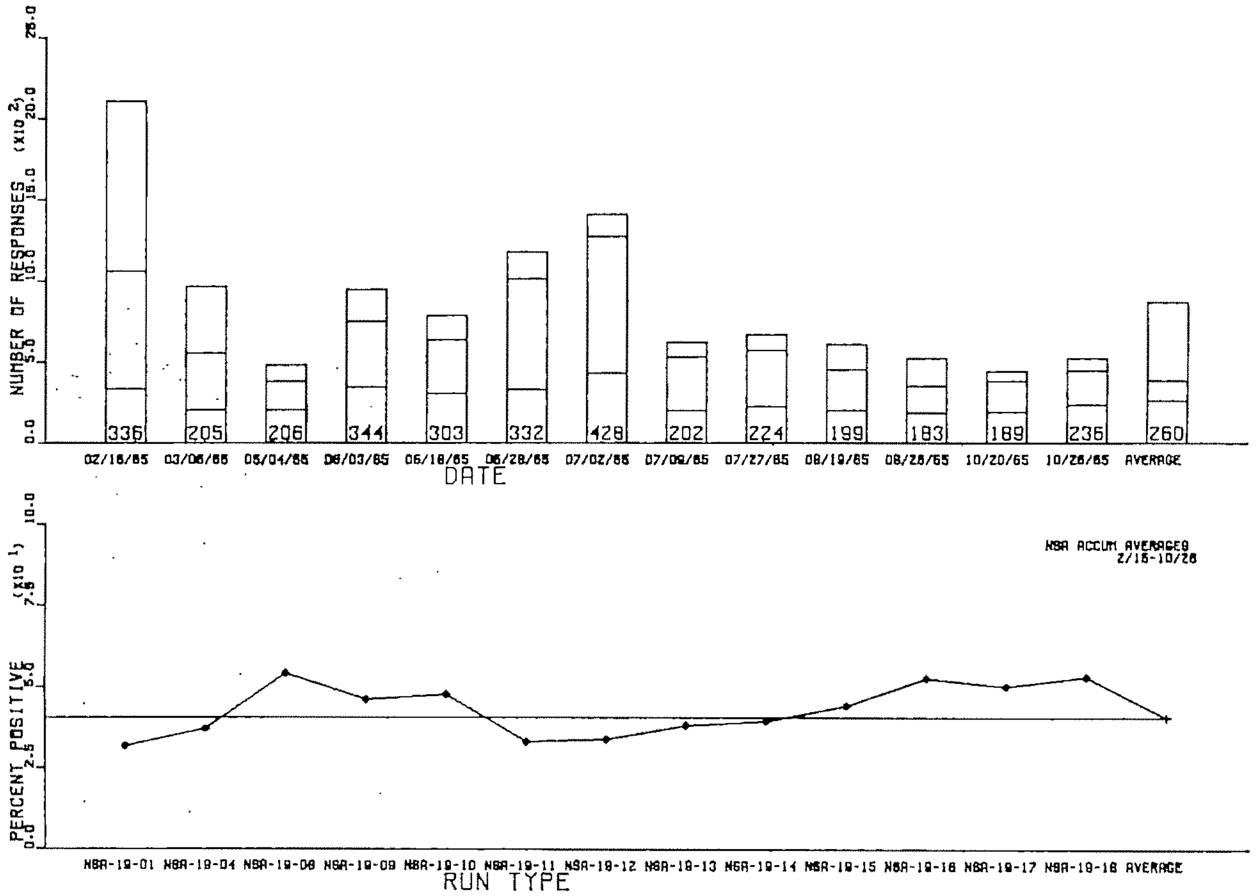


FIG. 3. *Nuclear Science Abstracts* accumulated averages for comparative study

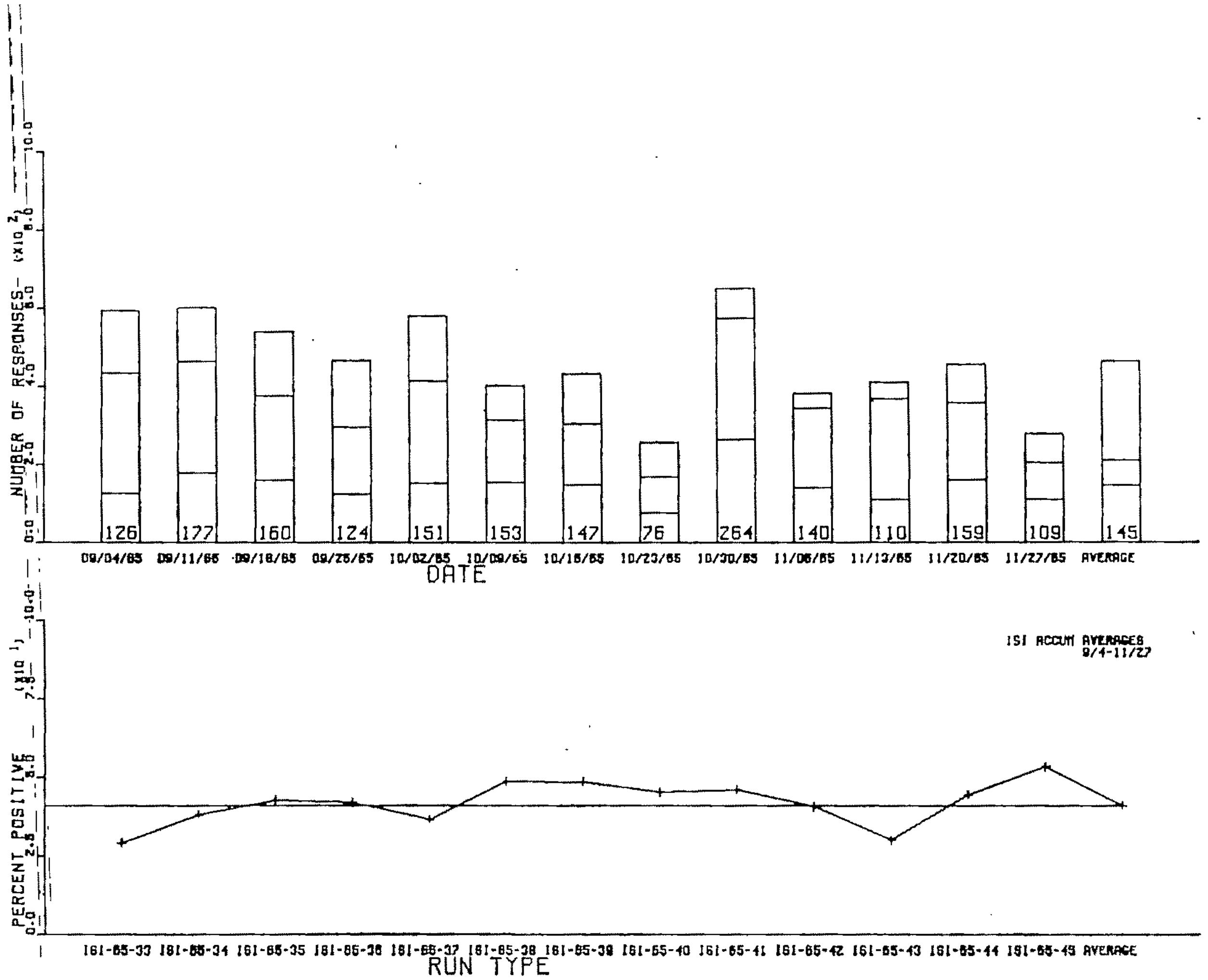


Fig. 4. *Science Citation Index* accumulated averages for comparative study

USER NAME	PROFILE NUMBER	PERCENT PLUS	AVERAGE PLUS	AVERAGE NEGATIVE	AVERAGE TOTAL
USER-A	2701	47.06	59.62	67.08	181.23
USER-B	10801	19.72	1.08	4.38	30.54
USER-C	12501	41.54	18.69	26.31	54.08
USER-D	14601	37.07	6.33	10.75	25.17
USER-E	16401	36.61	20.08	34.77	68.62
USER-F	18601	35.69	8.92	16.08	32.08
USER-G	20301	21.95	11.08	39.38	65.15
USER-H	25701	44.08	17.46	22.15	62.15
USER-I	31401	61.29	8.77	5.54	17.69
USER-J	31501	20.00	1.31	5.23	7.31
USER-K	33401	22.58	10.92	37.46	54.77
USER-L	33701	53.98	24.00	20.46	53.46
USER-M	36101	37.55	14.15	23.54	45.31
USER-N1	40101	44.32	9.31	11.69	25.77
USER-N2	40105	65.66	5.42	2.83	9.33
USER-N3	40107	43.72	12.85	16.54	33.46
USER-O	41001	67.97	12.08	5.69	22.77
USER-P	41601	23.08	5.08	16.92	22.15
USER-Q	76601	50.30	12.85	12.69	46.77
USER-R	86601	24.68	1.73	5.27	9.91

Fig. 5a. Accumulated averages of 13 runs (*Nuclear Science Abstracts*) from February 16, 1965, through October 26, 1965

USER NAME	PROFILE NUMBER	PERCENT PLUS	AVERAGE PLUS	AVERAGE NEGATIVE	AVERAGE TOTAL
USER-A	2701	37.73	34.08	56.23	150.77
USER-B	10801	21.05	1.85	6.92	11.38
USER-C	12501	56.02	16.46	12.92	31.08
USER-D	14601	74.36	2.23	0.77	4.54
USER-E	16401	69.03	6.00	2.69	14.54
USER-F	18601	35.98	5.92	10.54	29.69
USER-G	20301	47.89	2.62	2.85	8.00
USER-H	25701	29.36	5.31	12.77	18.46
USER-I	31400	55.17	1.23	1.00	2.62
USER-J	31501	47.73	1.62	1.77	4.23
USER-K	33401	9.46	0.54	5.15	6.00
USER-L	33701	43.04	13.08	17.31	30.62
USER-M	36101	50.76	5.15	5.00	11.54
USER-N1	40101	77.73	12.62	3.62	25.38
USER-N2	40105	33.42	10.00	19.92	30.46
USER-N3	40106	90.91	2.31	0.23	2.85
USER-N4	40107	70.00	10.77	4.62	16.54
USER-O	41001	75.00	2.54	0.85	4.08
USER-P	41601	44.89	7.77	9.54	18.62
USER-Q	76601	8.88	3.46	35.54	43.38
USER-R	86601	71.43	1.92	0.77	4.15

FIG. 5b. Accumulated averages of 13 runs (*Science Citation Index*) from September 4, 1965, through November 27, 1965

cance values which adapt to less descriptive source documents will be too sensitive to more descriptive source documents; the end result being a number of more descriptive source documents flooding users with document notifications having a low relative percentage of interest. Less descriptive source documents would be too discriminate in selection resulting in a high relative percentage of interest. Therefore, we lowered the threshold for the less descriptive source document, *Science Citation Index*, based on numeric ratio of words, intuition, and an "educated" guess. From the positive percentages derived, the relative positive percentages differed by only .61%.

To further verify our findings we decided to plot comparative percentages of positive responses based on individual percentage averages rather than on total number. Figure 6 shows the graphic results of these findings; the dotted line being the average of the two averages. The total averages differed here by only 2.17%, *Nuclear Science Abstracts* being 48.17% and *Science Citation Index* being 50.34%.

Other interesting figures to be noted were the close ratios of numeric quantities between source documents of the neutral, negative, and positive responses. Because of these ratios, some questions could be raised about the value of the keywording scheme used in *Nuclear Science Abstracts*. Another interesting figure was the differenc in the degree of selectivity, the average number of positive responses per user in relation to the average numbe of document entries per dissemination run. The degre of selectivity for *Nuclear Science Abstracts* was 2.1% an .36% for *Science Citation Index*. We consider this difference reasonable because of the differences in scientifi coverage of both source documents. This definitely indicates that the system is adapting not only to actual document composition but also to differentials in disciplin coverage, which is most encouraging.

• IV. Results of Ames Lab SDI, 1965

The sources of documents used over the 40-run perio described here were the same as described in Section III Only 13 *Nuclear Science Abstract* runs were made durin 1965 out of a possible 26 runs which might have bee made. This particular tape service is in only its pilo phases and many problems arise in its preparation whic result in incomplete usage of this source document. Al *Nuclear Science Abstracts* source documents were ru with a .5 threshold.

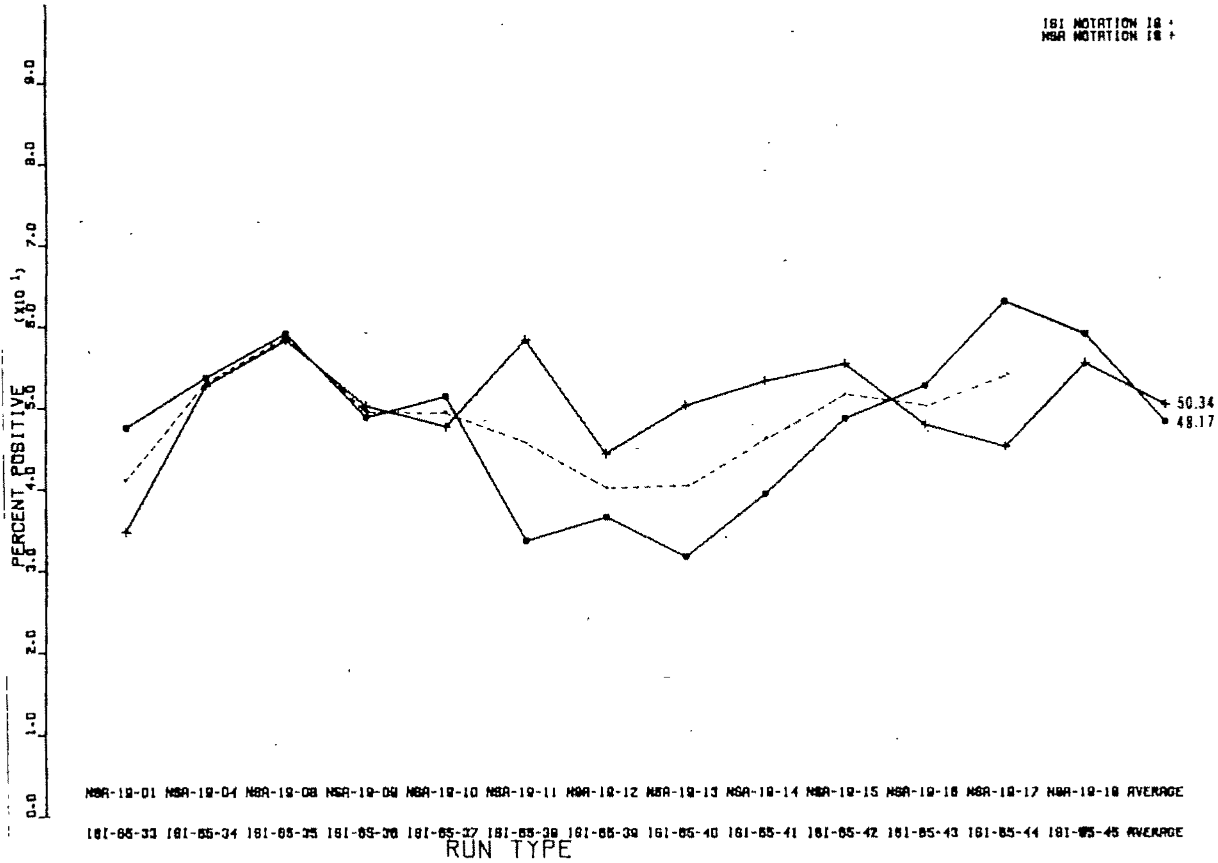


FIG. 6. *Science Citation Index* and *Nuclear Science Abstracts* relative positive percentage derived on individual averages for comparative study.

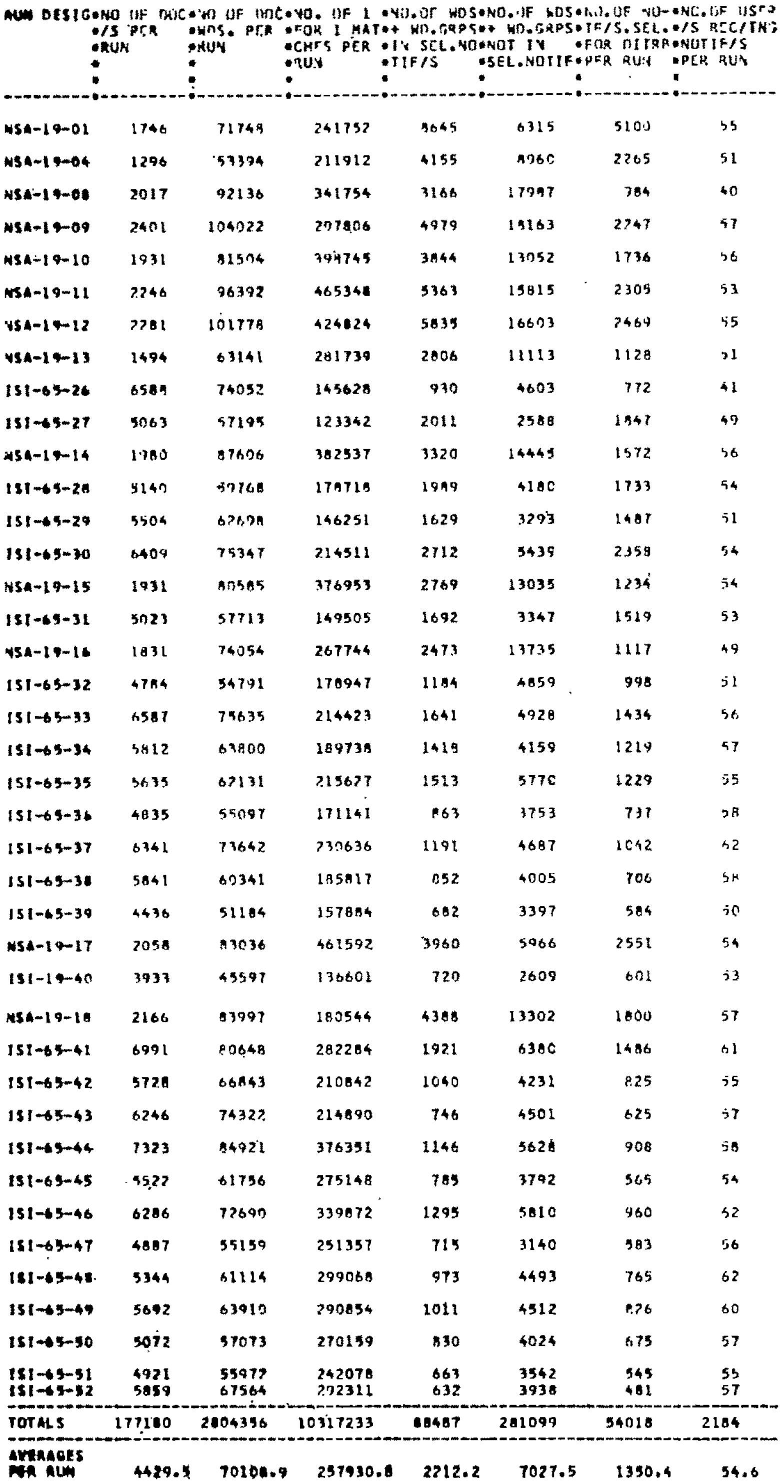
Science Citation Index source documents were run weekly for a six-month period from July 1965 to January 1966. A threshold value of .5 was used for the first three runs of *Science Citation Index* and .3 for the remaining 24 runs.

A total of 177,180 document entries was scanned through 40 runs involving 2,804,356 eligible terms for matching. Eight percent of the original number of terms were considered "general" and were automatically deleted in the scanning program. These 2,804,356 document terms yielded 10,317,233 matches with single profile word terms; 369,586 of these matches were used in the selection process resulting in 54,018 notifications selected for distribution to profile users. A breakdown of the number of document entries, document terms, matched terms, matched terms in selected documents, matched terms in nonselected documents, number of notifications selected, and number of users receiving notifications for each of the 40 runs is shown in Fig. 7. The category "matched terms in nonselected documents" refers to complete words or word groups matching in documents for which notifications were not generated. This does not include incomplete word groups or terms affected by negation.

Figure 8 (a and b) shows an analysis of the word structuring of all profiles as they appeared after the final run for 1965. The majority of these profiles had undergone extensive revisions during the 40-run period and many hardly resembled their original state. We had an average of 3.6 manual revisions of profiles per week. These included the addition and deletion of words and word groups, readjustment of significance values, and, in some cases, complete rewording and restructuring. We did not keep an accurate account of these revisions and cannot reflect this variable in the number of notifications distributed or percentages of relevancy.

Figures 9–16 present the graphic results of the number of notifications disseminated, represented by bar graphs indicating the number of neutral, negative, and positive responses. Below each bar is the number of users who received notifications for that particular run. The line graphs are the relative percentages of positive notifications, the solid line represents the total number of positive and negative notifications for each run and the dotted line indicates the average individual user percentages for each run. The dotted and solid straight lines running across the line graph are the corresponding total percentage averages for the 40-run period. The numeric averages are also printed in the line graph.

Seven groups were chosen and the profile users were categorized according to their research interests into one



RUN DESIG	NO OF DOC /S PER RUN	NO OF DOC WDS. PER RUN	NO. OF 1 FOR 1 MAT CHES PER RUN	NO.OF WDS + WD.GRPS IN SEL.NO TIF/S	NO.OF WDS + WD.GRPS NOT IN SEL.NOTIF	NO.OF NO- TIF/S.SEL. FOR DISTR PER RUN	NO.OF USER /S REC/ING NOTIF/S PER RUN
NSA-19-01	1746	71748	241752	3645	6315	5100	55
NSA-19-04	1296	53394	211912	4155	8960	2265	51
NSA-19-08	2017	92136	341754	3166	17987	784	40
NSA-19-09	2401	104022	297806	4979	15163	2747	57
NSA-19-10	1931	81504	398745	3844	13052	1736	56
NSA-19-11	2246	96392	465348	5363	15815	2305	53
NSA-19-12	2281	101778	424824	5835	16603	2469	55
NSA-19-13	1494	63141	281739	2806	11113	1128	51
ISI-65-26	6585	74052	145628	930	4603	772	41
ISI-65-27	5063	57195	123342	2011	2588	1847	49
NSA-19-14	1980	87606	382537	3320	14445	1572	56
ISI-65-28	5140	59768	178718	1989	4180	1733	54
ISI-65-29	5504	62698	146251	1629	3293	1487	51
ISI-65-30	6409	75347	214511	2712	5439	2358	54
NSA-19-15	1931	80585	376953	2769	13035	1234	54
ISI-65-31	5023	57713	149505	1692	3347	1519	53
NSA-19-16	1831	74054	267744	2473	13735	1117	49
ISI-65-32	4784	54791	178947	1184	4859	998	51
ISI-65-33	6587	75635	214423	1641	4928	1434	56
ISI-65-34	5812	63800	189738	1418	4159	1219	57
ISI-65-35	5635	67131	215627	1513	5770	1229	55
ISI-65-36	4835	55097	171141	863	3753	737	58
ISI-65-37	6341	73642	230636	1191	4687	1042	62
ISI-65-38	5841	60341	185817	052	4005	706	58
ISI-65-39	4436	51184	157884	682	3397	584	50
NSA-19-17	2058	83036	461592	3960	5966	2551	54
ISI-19-40	3933	45597	136601	720	2609	601	53
NSA-19-18	2166	83997	180544	4388	13302	1800	57
ISI-65-41	6991	80648	282284	1921	6380	1486	61
ISI-65-42	5728	66843	210842	1040	4231	825	55
ISI-65-43	6246	74322	214890	746	4501	625	57
ISI-65-44	7323	84921	376351	1146	5628	908	58
ISI-65-45	5522	61756	275148	785	3792	565	54
ISI-65-46	6286	72690	339872	1295	5810	960	62
ISI-65-47	4887	55159	251357	715	3140	583	56
ISI-65-48	5344	61114	299068	973	4493	765	62
ISI-65-49	5692	63910	290854	1011	4512	826	60
ISI-65-50	5072	57073	270159	830	4024	675	57
ISI-65-51	4921	55977	242078	663	3542	545	55
ISI-65-52	5859	67564	292311	632	3938	481	57
TOTALS	177180	2804356	10317233	88487	281099	54018	2184
AVERAGES PER RUN	4429.5	70108.9	257930.8	2212.2	7027.5	1350.4	54.6

FIG. 7. Statistical results of 40 dissemination runs

Metallurgy

USER/S NAME	PROFILE NO.	NU.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NU.5 WDS.		NU.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	PUS.	NEG.
METALLURGY-A	12501	44	28	101		11		3					
METALLURGY-B	3701	7	14	1	9								
METALLURGY-C	18601	77	80	85									
METALLURGY-D1	20002	2		3		1		3		2			
METALLURGY-D2	20003			2		1							
METALLURGY-D3	20004	1		2		2		4		2			
METALLURGY-D4	20005	4		3		6							
METALLURGY-D5	20006			1		6							
METALLURGY-E	20301	49	4	35		1							
METALLURGY-F	20501	29		24									
METALLURGY-G	31501	3	13	48		64		7		2			
METALLURGY-H	76601	136	75	29	22	34	13	29	6	4			
TOTALS-----NO.USERS	12	345	200	333	22	126	13	46	6	10			

Chemistry

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
CHEMISTRY-A	2701	11	117	1042	65	9	6						
CHEMISTRY-B	7201	8	3	13		2		2					
CHEMISTRY-C	7401	32	4	64		2							
CHEMISTRY-D	9101	9		15		14							
CHEMISTRY-E1	14601			81		15		1					
CHEMISTRY-E2	14602			5		3		1					
CHEMISTRY-F	31501	3	13	48		64		7		2			
CHEMISTRY-G	25701	47		65		3	10						
CHEMISTRY-H	32501	5		2		12							
CHEMISTRY-I1	40101	8	1	166		3							
CHEMISTRY-I2	40102	1	1	1									
CHEMISTRY-I3	40103	4		23		3							
CHEMISTRY-I4	40104	2		11									
CHEMISTRY-I5	40105	5		7		6							
CHEMISTRY-I6	40106			12									
CHEMISTRY-I7	40107	12		20		7		1		1			
CHEMISTRY-I8	40109			13									
CHEMISTRY-J	54001		1	45		25		8		2			
CHEMISTRY-K	56101	64	345	104	12	6	1						
CHEMISTRY-L1	74501	24		16		6							
CHEMISTRY-L2	74502	2		2									
CHEMISTRY-L3	74503	2		5		4		3					
CHEMISTRY-L4	74504	5	1	12	1	6		7		1			
CHEMISTRY-L5	74505			2		6							
CHEMISTRY-L6	74506	16		10		1		1					
CHEMISTRY-L7	74507		3	1		17		2					
CHEMISTRY-M	84301	5	5	11	1	4		2					
CHEMISTRY-N1	86601	2		15		4							
CHEMISTRY-N2	86602	27		44		9							
CHEMISTRY-N3	86603	11		68		2							
TOTALS-----NU.USERS	30	305	494	1923	79	233	17	35		6			

FIG. 8a. Analysis of all Ames Lab SDI Profiles

Experimental Physics

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
EXP.PHYS-A	31501	3	13	48		64		7		2			
EXP.PHYS-B	33701	41	16	54									
EXP.PHYS-C	35201	1		11		5							
EXP.PHYS-D	36101	24	27	101	5	10		1					
EXP.PHYS-E	38701	30	2	11									
EXP.PHYS-F	39201	14	28	7	1	2		2					
EXP.PHYS-G1	40301	3	8	39		1							
EXP.PHYS-G2	40302	12	7	17									
EXP.PHYS-G3	40303	27	7	30									
EXP.PHYS-H	41001	9		18		7		1					
EXP.PHYS-I	41401	17	8	68		5	2						
EXP.PHYS-J	41501	14	1	139		5		8					
EXP.PHYS-K	41601	105	1	52		19		2					
EXP.PHYS-L	42001	3	10	29	1	5		1					
EXP.PHYS-M	45801	1		9		8							
EXP.PHYS-N1	47601	6		25		5							
EXP.PHYS-N2	47602	47		61									
EXP.PHYS-N3	47603	48		71									
EXP.PHYS-O	78501	20	31	44	1	3							
TOTALS-----NO.USERS	19	411	158	695	8	134	3	14		2			

Theoretical Physics

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
THEOR.PHYS-A	33401	34	1	26		5							
THEOR.PHYS-B	33501	2		23		2		1					
THEOR.PHYS-C	33601	50	3	23		3							
THEOR.PHYS-D	40201	1		13		12		6					
THEOR.PHYS-E	34801	16	5	102	1	9							
TOTALS-----NO.USERS	5	103	9	187	1	33		7					

Engineering

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
ENGINEERING-A	10801	28	6	68		6	2						
ENGINEERING-B	16401	20	1	25	3	5		2					
ENGINEERING-C	31400	15		19		10		3					
ENGINEERING-D	53401	16	4	19		4							
ENGINEERING-F	79801	32	12	37		6							
TOTALS-----NO.USERS	5	111	23	168	3	31	2	5					

Reactor

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
REACTOR-A1	50401	2		13		34		5					
REACTOR-A2	50402	8		13		23		2					
REACTOR-A3	50403	2		23		11		4					
REACTOR-A4	50404			2		7							
REACTOR-A5	50405			3		4							
REACTOR-A6	50406			4		2							
TOTALS-----NO.USERS	6	12		58		81		11					

Mathematics and Computers

USER/S NAME	PROFILE NO.	NO.1 WD.		NO.2 WDS.		NO.3 WDS.		NO.4 WDS.		NO.5 WDS.		NO.6 WDS.	
		POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.	POS.	NEG.
MATH.COMP-A	9701	11		25		8		1					
MATH.COMP-B	88101	17		39		1							
MATH.COMP-C	87901	12		46									
MATH.COMP-D	92301	6		13									
TOTALS-----NO.USERS	4	46		123		9		1					

FIG. 8b. (*Continued*)

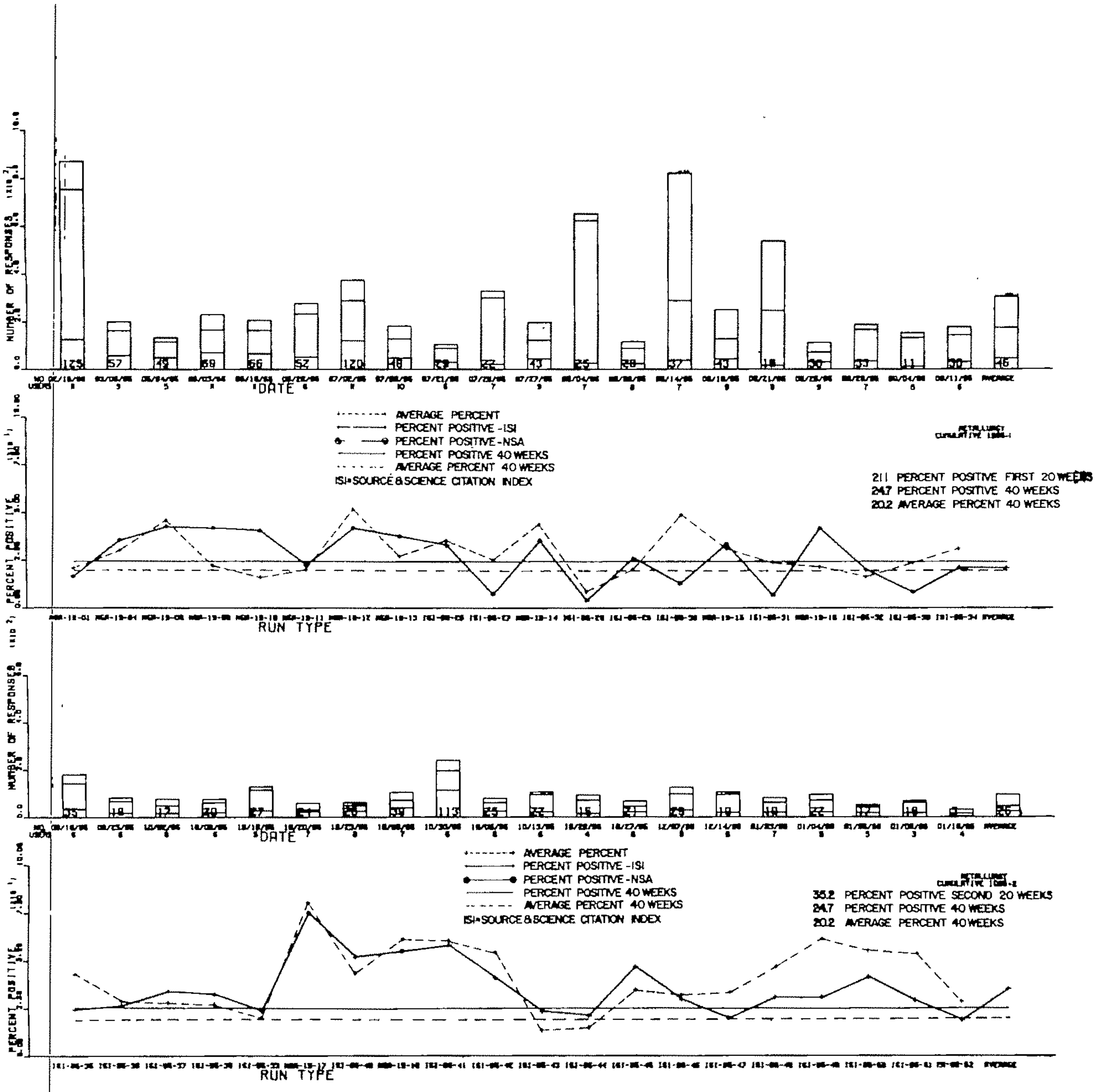


FIG. 9. Accumulated averages of metallurgy group for 40 runs

or more groups. Each individual is involved in a specific area of research within each group. However, we thought some insights could be gained with the loose categorized grouping into scientific disciplines. A detailed study of individual profile performance within groups would reveal diverse results from SDI, but to graphically represent these results would be too lengthy for this article. Test profiles and their selected notifications were not shown in the graphic results, although their totals are included in Fig. 7. We had no means at our disposal by which to screen these totals, but they did not significantly distort the figures presented.

Figure 9 contains the graphic results of 12 scientists representing the field of metallurgy. This particular group achieved the poorest relative percentages of the seven groups represented. There was a great variance in the number of notifications and relative percentages from one run to the next. Approximately 40% of the users' profiles did not receive notifications for a given run. Many complaints were filed from this group concerning weak subject coverage of both source documents. The general construction of metallurgy profiles was not as extensive as the average profile of all groups. A unique factor with this group was the inclusion of many weighted positive journal source notations which resulted in users receiving all articles within a given journal. The smoothing of the percentage lines of the last 10 runs was caused by journal source notation significance values having been decremented heavily from feedback of earlier runs.

The chemistry group possessed the largest number of users and the most comprehensive profiles. In Fig. 8a it appears that single users have many profiles when actually these profiles represent junior scientists and graduate students working under a senior scientist. The

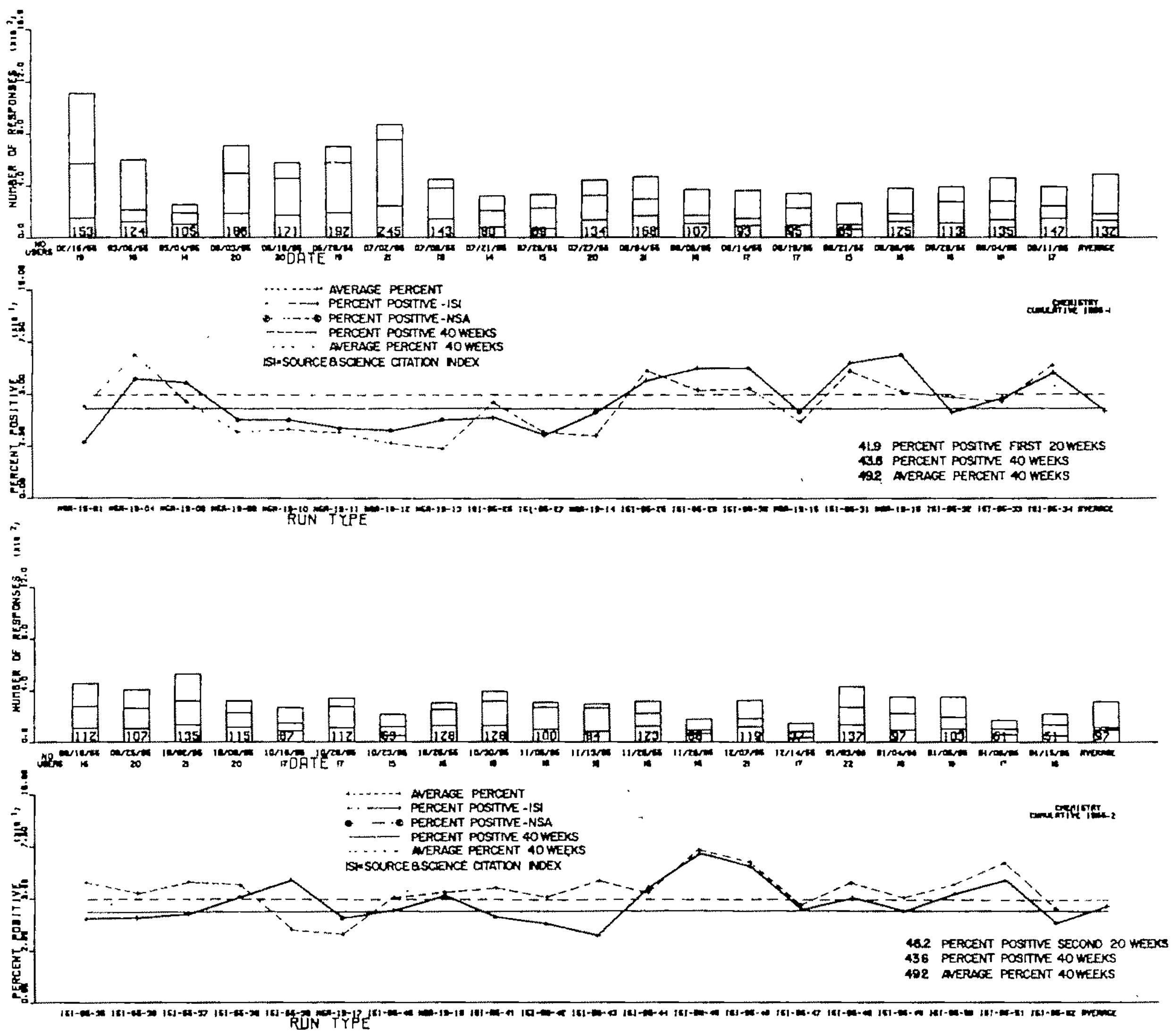


FIG. 10. Accumulated averages of chemistry group for 40 runs

numbers of notifications selected and relative positive percentages were evenly distributed over the 40-run period considering the normal fluctuations of journal distribution and the addition of 14 new profiles. The people in this group did not have any complaints about weak coverage of subject matter, and the majority felt their current literature coverage had been expanded because of SDI. The nine small profiles of this group received very few notifications after their initial runs and had very little effect on the results of this group except in the calculation of positive percentages based on individual results.

The third group, experimental physics, was close to the norm in terms of profile construction and results obtained. Initially, problems arose in the system in adjusting to *Nuclear Science Abstracts* and *Science Citation Index* source documents. The first 20 runs, as noted in Fig. 11, indicated poor results. Consequently, the users of this group revamped their profiles extensively, adding a significant number of authors and negative journal sources. In this group, authors accounted for 53.8% of the positive profile terms, and 89.2% of the negative terms were journal sources. The individuals of this group received the most notifications per user. Two scientists of this group felt the literature coverage of the two source documents was weak, while the remaining users considered the coverage to be adequate. The members of this group seemed particularly anxious to be notified of the latest journal articles and had very little use for articles published six or more months previously.

Figure 12 represents the graphic results of the theoretical physics group. Figure 8b shows there were only five profiles in this group which resulted in dramatic fluctuations in the relative percentage lines. One of the five users showed little enthusiasm for SDI and received very few notifications. His disinterest affected the relative percentage of individual averages. The users of this group did not make any manual revisions to their profiles after they submitted their originals. There was an indication of improvement with the number of notifications disseminated and relative positive percentages during the last 15 runs. This improvement was attributed to the feedback mechanism.

The engineering group also had only five profiles, reflecting diverse engineering interests. These interests included chemical, ceramic, and mechanical engineering. The first 20 runs, Fig. 13, produced lower results than the second 20 runs. However, it seems obvious from the number of notifications disseminated that the bulk of pertinent scientific literature for engineering was selected from *Nuclear Science Abstracts* source documents. Parallel with the theoretical physics group few manual revisions were made to engineering profiles and the advent of *Science Citation Index* source documents indicated that these profiles were not extensive enough to select documents on title, author, and source information only.

Figure 14 shows the graphic results of the reactor group. The performance of the reactor group profiles, although small in number, indicated obvious discrepancies in subject coverage of the source documents. *Nuclear Science Abstracts,* as an abstracting journal, is adequate in servicing this group with perhaps the periodic selection of fringe literature from *Science Citation Index* source documents. Because of the small number of notifications selected for these users over the 40-run period, it was doubtful if the significance values of the words and word groups reached equilibrium after the last run. Much of the literature covering reactor technology is generated by USAEC scientists and published in R&D report form. This indicates that *Nuclear Science Abstracts* source documents are more apropos to this particular group than *Science Citation Index* source documents.

The last group, mathematics and computers, was the one biased group of the seven. Three of the users are directly involved with the Ames Laboratory SDI system, one the author of this article. From a personal standpoint, the notifications disseminated from *Science Citation Index* source documents were quite pertinent to his interests, particularly articles from noncomputer oriented journals. Eight articles were selected during this 40-run period describing SDI systems, three of these articles were from noncomputer, nondocumentation journals. The graphic results of this group are represented in Fig. 15. It should be noted from the graphs that very few notifications were disseminated from the *Nuclear Science Abstracts* source documents; four of the users negated the complete source.

Figure 16 shows the composite graphic results of the seven groups. The relative positive percentages were somewhat erratic during the first 20 runs, but definitely smoothed out during the second 20 runs. Feedback was the major contributor to this improvement with manual profile revisions contributing as the second factor. The percentages may seem low with the published results of other SDI systems; however, one must be aware that 100% of the notifications selected and disseminated were used in the calculations of these percentages. The responses to notifications returned for feedback from a given run were not, in most cases, received prior to the next run. The average cut-off period (all notifications prior to the cut-off date were treated as negative responses) was 4.4 weeks from the date the notifications were distributed. The average degree of selectivity (the percentage of document entries of notifications out of the total number of document entries scanned) per user for the entire 40-run period was 0.55%. Considering the diverse subject coverage of our source documents, we were quite elated over this percentage even though we are not aware of pertinent documents not being selected for dissemination.

The number of users representing four groups—theoretical physics, engineering, reactor, and mathematics and computers—was small. Unfortunately we had no control over this situation. It was apparent one would require a minimum of 10 to 15 users per group to measure the comprehensiveness of source documents related to a particular scientific discipline. Table 1 represents the average percentages and numbers of selected notifications as represented graphically in Figs. 9–16.

• Conclusion

There are various areas within our SDI system which cause "noise" and inefficiencies. Utilizing two sources of input has created a burden for our users in designing their profiles. *Science Citation Index* follows various computer-oriented standards for authors, sources, and scientific notation while *Nuclear Science Abstracts* is more inclined to follow library-type standards. This situation has forced our users to define words and word groups of a particular research interest into various word combinations (e.g., C R Sage, C Sage, Charles R Sage, Charles Sage, Charles Russell Sage, Sage) to fully adjust to these input differentials. As mentioned earlier in this article, 43% of our users have requested that they be excluded from *Nuclear Science Abstracts* coverage which undoubtedly is due partially to the ambiguity of author and source notation and keywording. Certain editing routines will be incorporated into the system where algorithms and reasonably short table "look-ups" can compensate for some of these ambiguities. Having incorporated six various sources of input into this system we see a very critical need for universal standards. A subtle "noise" has been detected in matching authors and their respective initials as a two word group. With large volumes of documents the names normally considered unique are not unique and this produces many irrelevant notifications. This factor is one answer for the large number of neutral responses. Cross matches are occurring with initials and

Table 1. Accumulated Group Averages of SDI Responses.

Name	Profile	Percent plus	Average plus	Average negative	Average total
Chemistry	Cumulative 1965—1st Half	41.94	132.50	183.40	489.25
Chemistry	Cumulative 1965—2nd Half	46.17	97.85	114.10	316.10
Chemistry	Cumulative 1965—Total	43.64	115.18	148.75	402.68
Engineering	Cumulative 1965—1st Half	33.42	47.55	94.75	177.15
Engineering	Cumulative 1965—2nd Half	42.62	13.85	18.65	45.10
Engineering	Cumulative 1965—Total	35.13	30.70	56.70	111.13
Reactor	Cumulative 1965—1st Half	51.67	22.84	21.37	58.58
Reactor	Cumulative 1965—2nd Half	28.47	10.89	27.37	42.37
Reactor	Cumulative 1965—Total	40.91	16.87	24.37	50.47
Metallurgy	Cumulative 1965—1st Half	21.10	46.40	173.50	304.75
Metallurgy	Cumulative 1965—2nd Half	35.15	26.45	48.80	95.95
Metallurgy	Cumulative 1965—Total	24.68	36.43	111.15	200.35
Computer	Cumulative 1965—1st Half	66.25	3.12	1.59	8.29
Computer	Cumulative 1965—2nd Half	47.31	10.35	11.53	26.47
Computer	Cumulative 1965—Total	50.66	6.74	6.56	17.38
Experimental physics	Cumulative 1965—1st Half	34.22	78.80	151.45	259.75
Experimental physics	Cumulative 1965—2nd Half	40.69	93.10	135.70	244.75
Experimental physics	Cumulative 1965—Total	37.45	85.95	143.58	252.25
Theoretical physics	Cumulative 1965—1st Half	20.83	12.55	47.70	73.20
Theoretical physics	Cumulative 1965—2nd Half	32.57	9.30	19.25	32.95
Theoretical physics	Cumulative 1965—Total	24.61	10.93	33.48	53.08

last names between co-authors of a particular document; a problem we did not forecast in designing the system. Compensations have been made in our "third generation" computer SDI system for processing author's last name and initials as a single term.

The peak input load for sorting efficiency on our present computer system using about 100 profiles is approximately 15,000 *Science Citation Index* or 5,000 *Nuclear Science Abstracts* entries per run. The running time of the computer asymptotically increases with numbers of additional documents greater than the aforementioned quantities. We have not calculated the maximum number of profiles for ideal running efficiency per run.

With large volumes of document entries which we process, there are still gaps in literature coverage of certain research interests. It is most important that these gaps be filled if we expect to have a comprehensive current awareness service. In the near future Iowa State University contemplates making this service available to its scientifically oriented faculty and at that time one would have a better knowledge of weak coverage over many diversified scientific disciplines (6).

The mathematical equations used in the feedback process are not the ultimate function in this application. We feel that the present equations are fulfilling our needs but there is room here for more experimentation. Investigations will be made to possibly refine this process when time permits. However, we are totally convinced that the concept of feedback is mandatory for effective comprehensive scientific information dissemination.

The bulk of input selected and disseminated to our users is composed of title, author, and source records. One might consider this document composition a low order document definition, yet the SDI system is adaptable enough to utilize these document entries. If an installation intends to institute an SDI system using relatively small numbers of documents composed of title, author, and source records, we recommend that they be KWOC (Keyword Out of Context) or KWIC (Keyword In Context) indexed and manually selected by users rather than by computer selection. From our experience with *Nuclear Science Abstracts,* keywording of documents is expensive and impractical for the relative additional benefits derived from SDI, particularly if it restricts numbers and coverage of documents. The utilization of author-derived abstracts seems to be the middle step for third generation computer systems. If authors do not write abstracts, one has another expensive problem. Ultimately, texts or partial texts will be the ideal form of input for adaptive SDI. We are running limited experiments with texts and abstracts, and in the early stages of experimentation, our present SDI system can hypothetically adjust and perform better with this form of input. Hopefully, fourth- and fifth-generation computers coupled with effective optical scanning devices will make this economically feasible. We are side stepping the development of automatic abstracting and keywording techniques because of limited resources and our own projection that necessary hardware will be available for text processing, thus making these theoretically developed techniques obsolete.

Testing and experimentation of an improved SDI system for an IBM 360/50 computer is underway. Additional features are being added to the system described in this article to include capabilities of generating quarterly for each individual profile a KWOC index of "OF

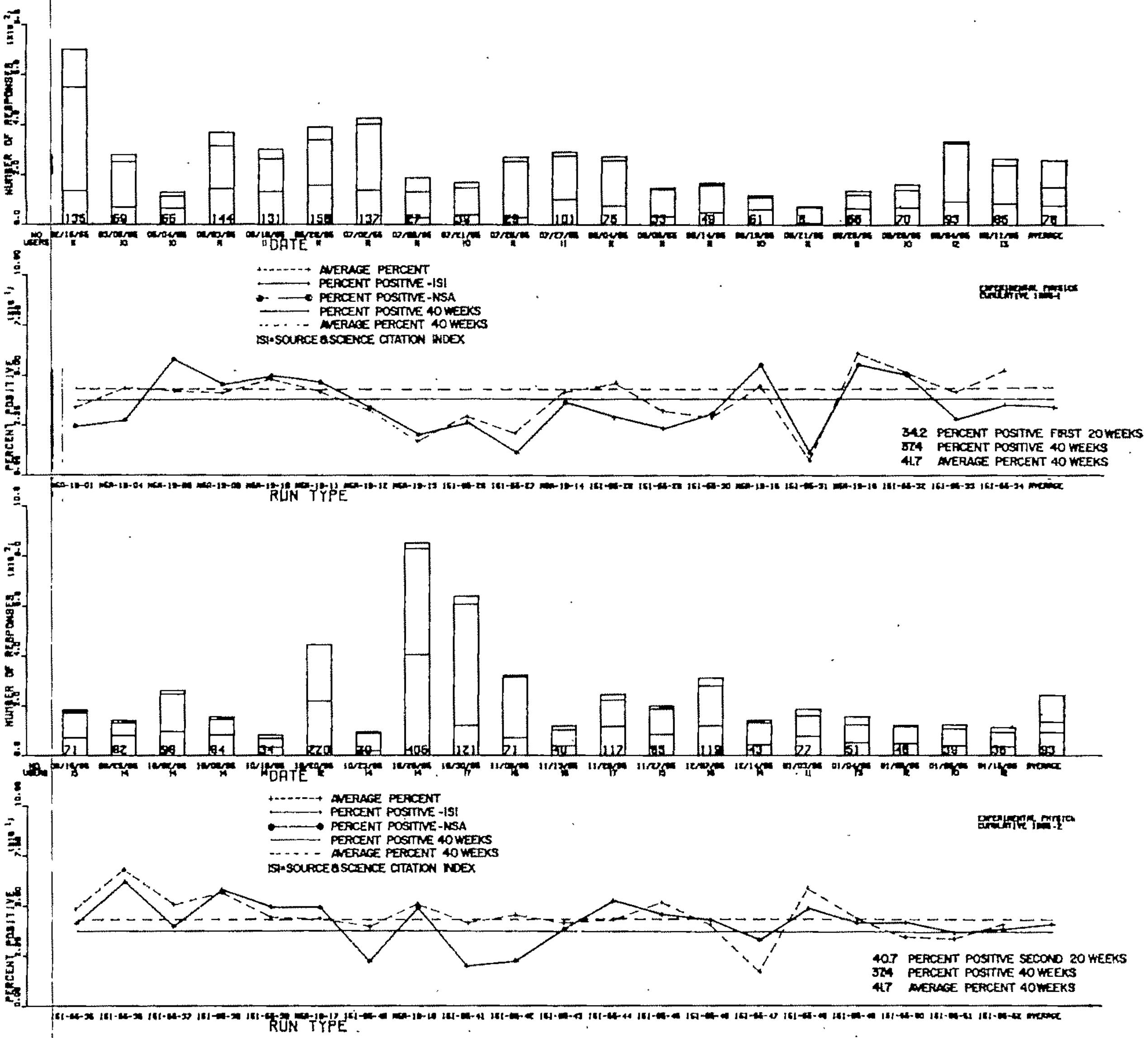


FIG. 11. Accumulated averages of experimental physics group for 40 runs

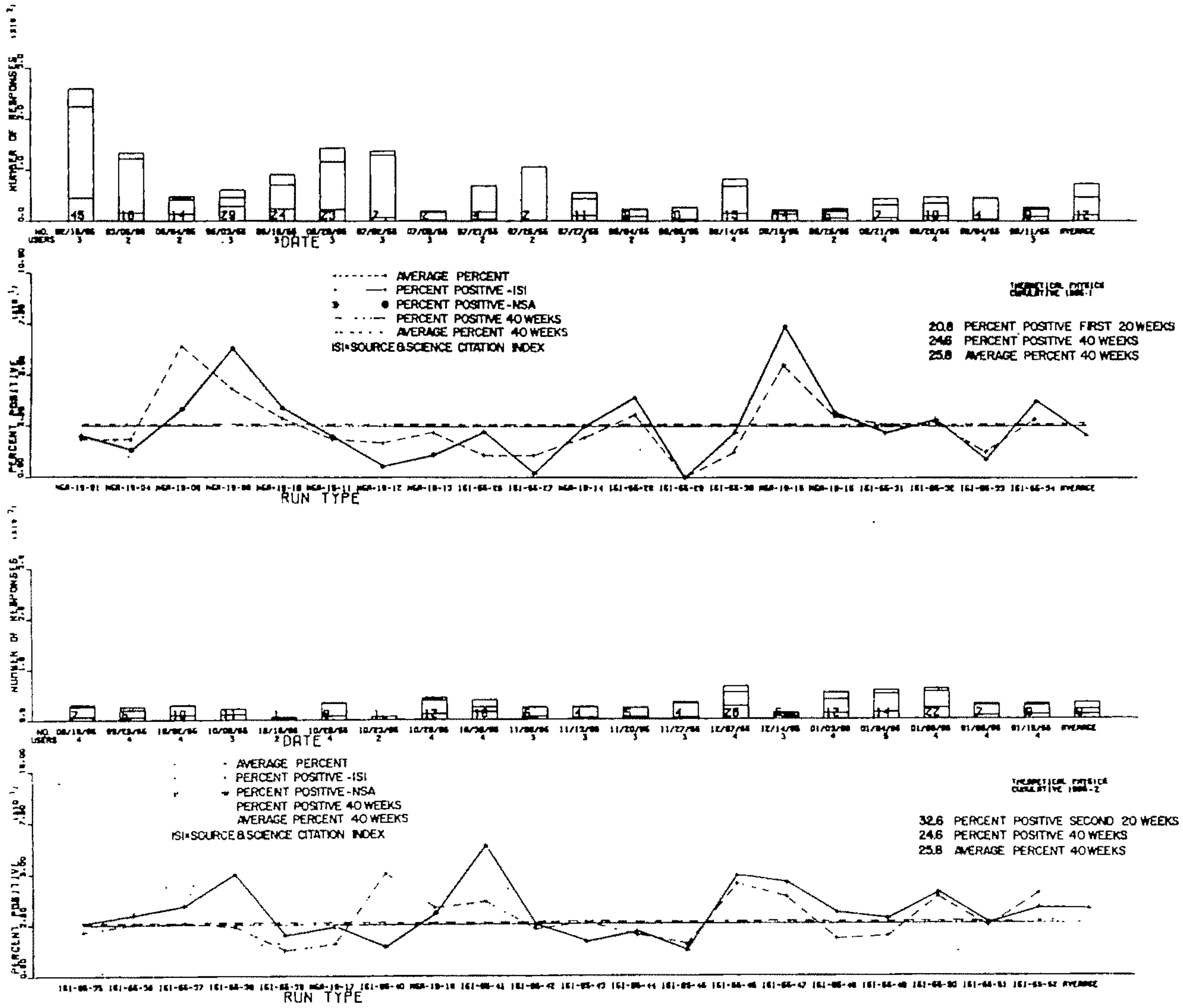


FIG. 12. Accumulated averages of theoretical physics group for 40 runs

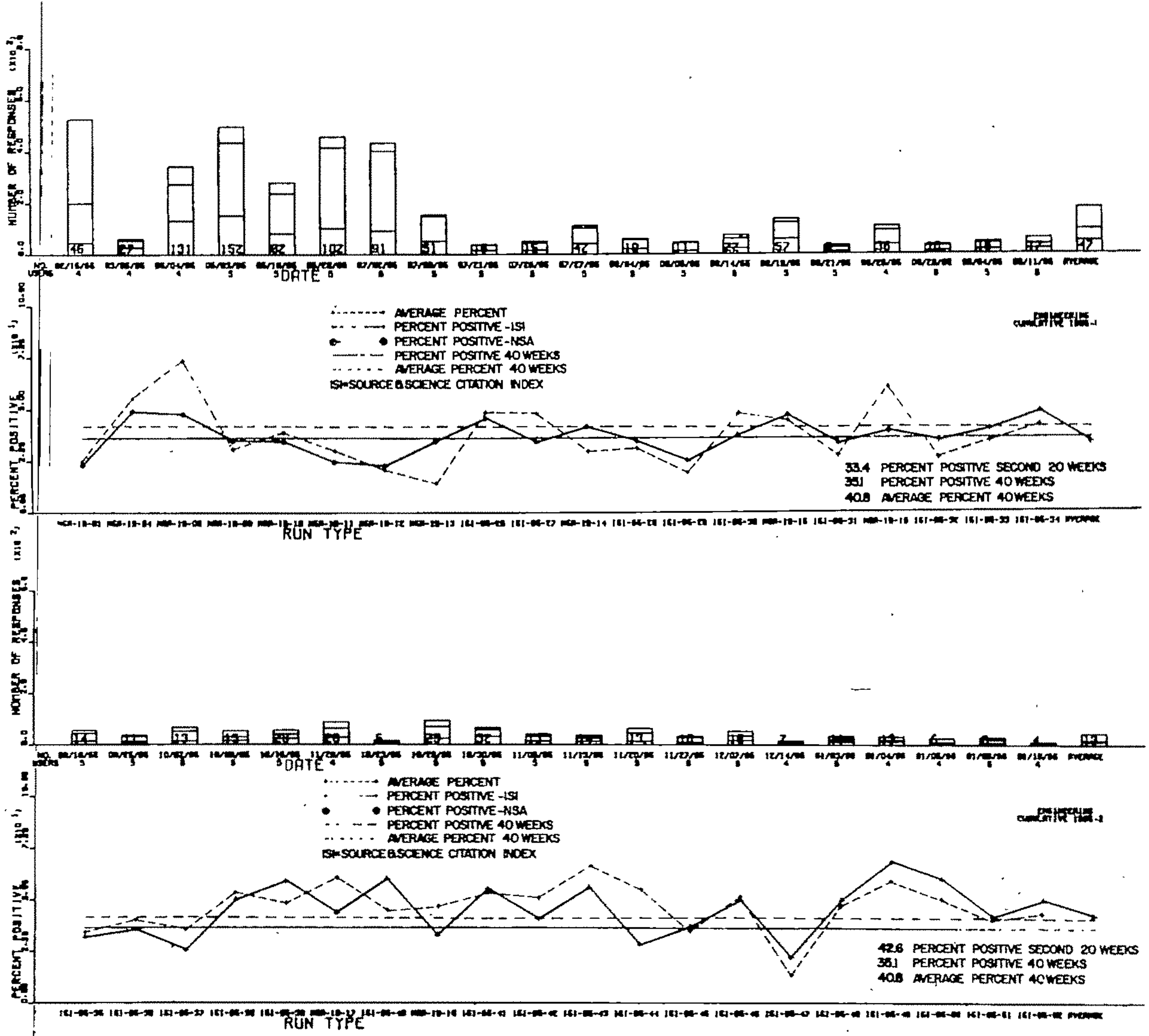


FIG. 13. Accumulated averages of engineering group for 40 runs

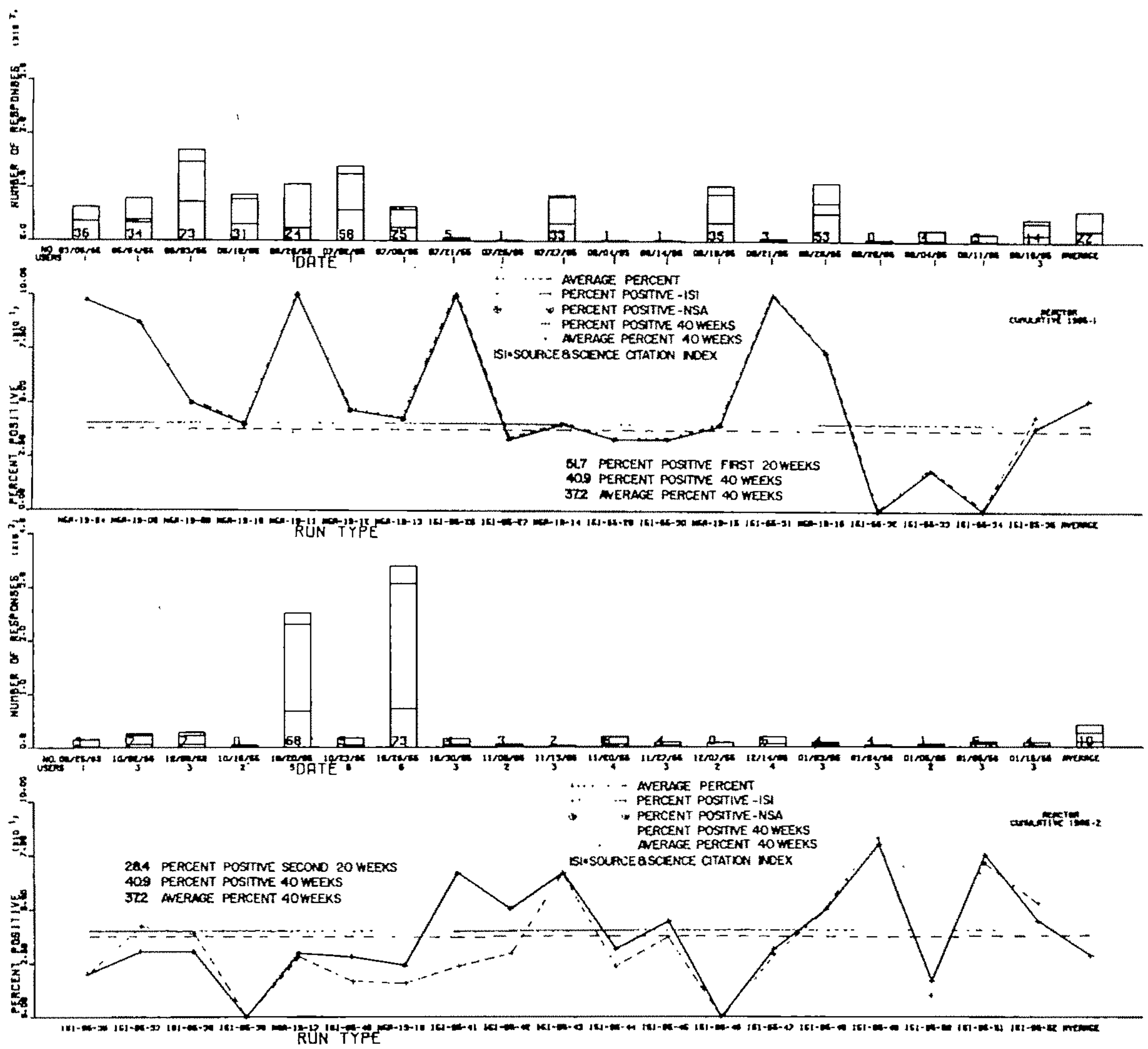


Fig. 14. Accumulated averages of reactor group for 40 runs

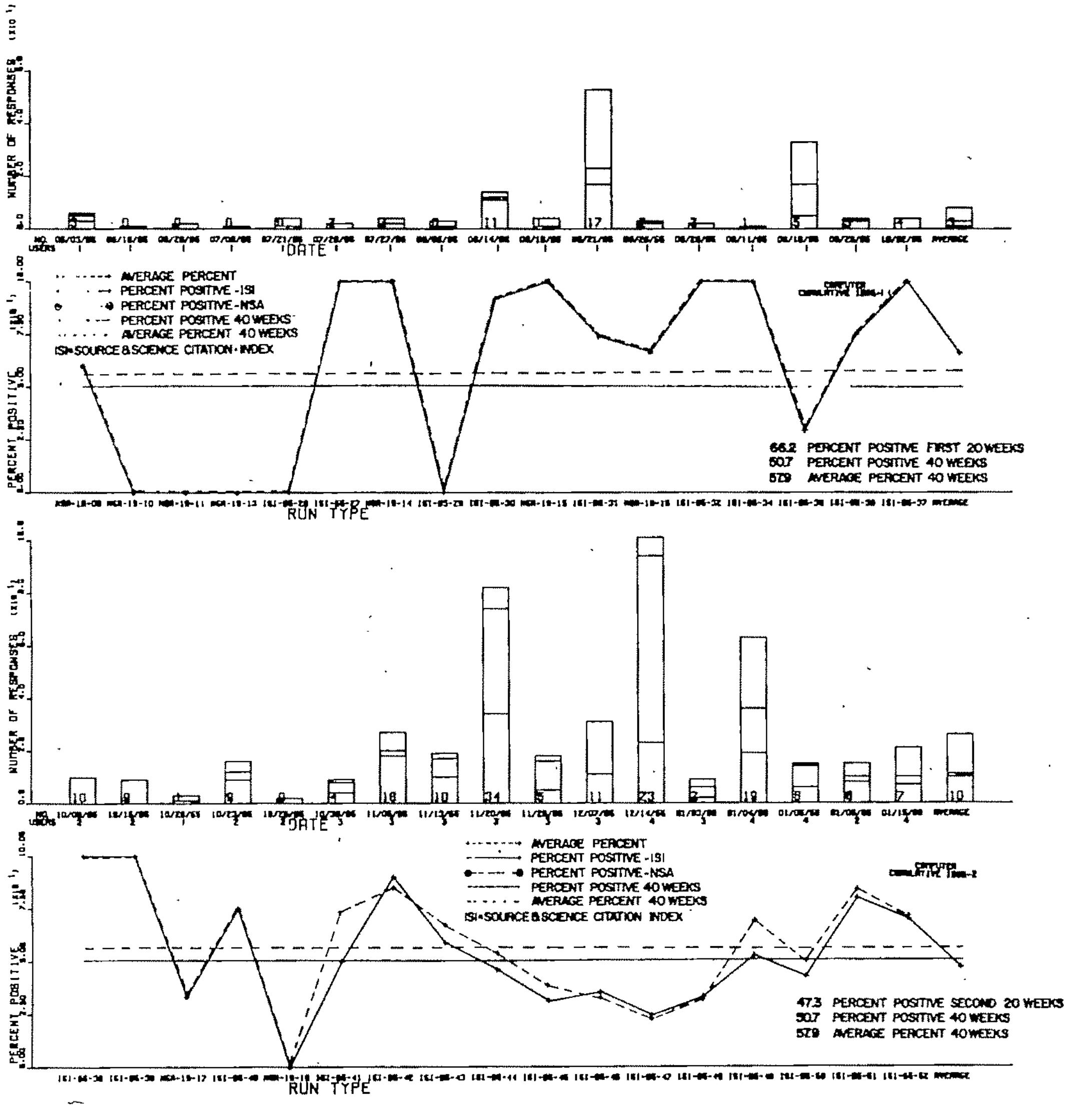


FIG. 15. Accumulated averages of mathematics and computers group for 40 runs

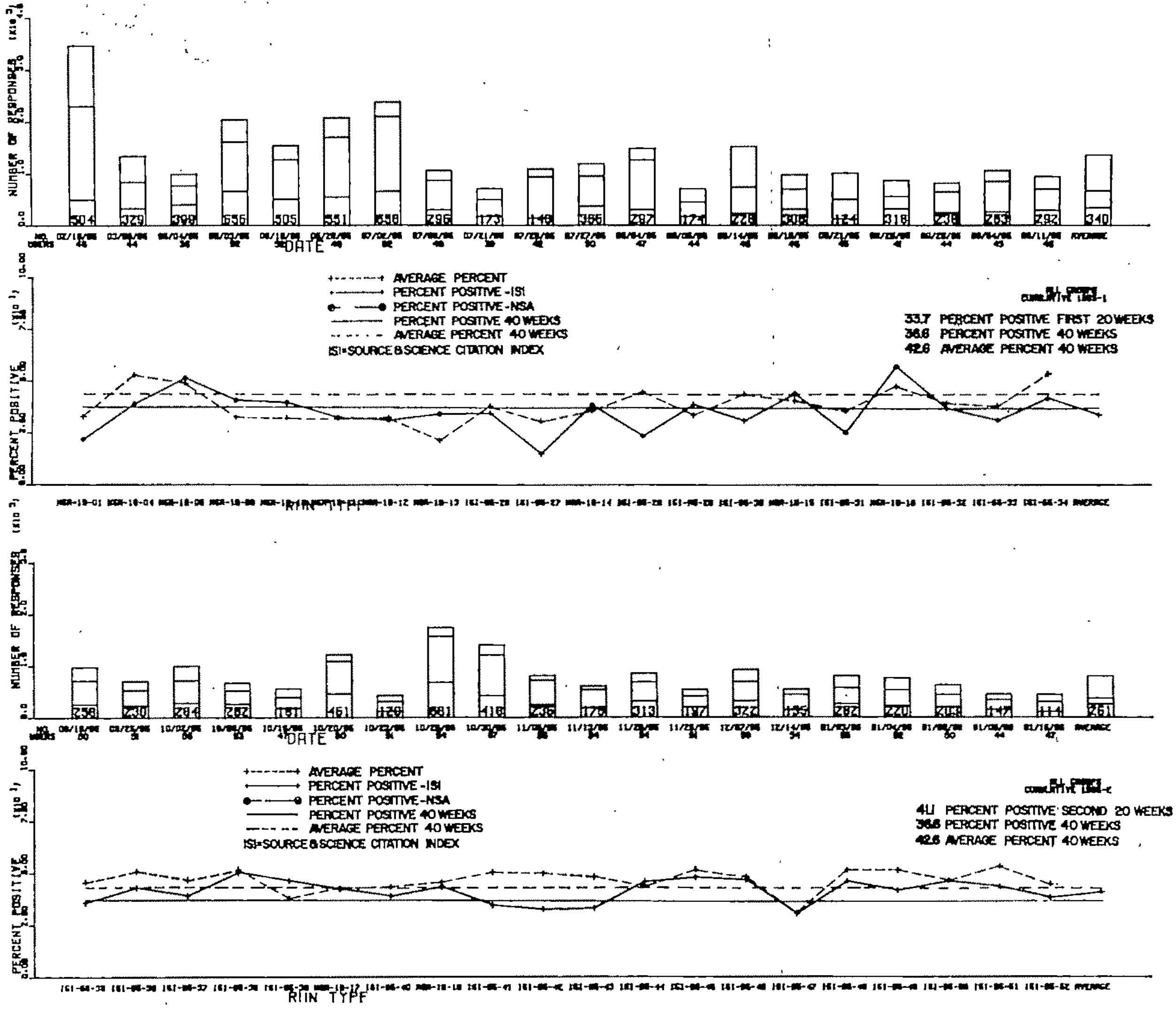


FIG. 16. Accumulated averages of all groups for 40 runs

INTEREST" articles which were returned during the quarter. Producing this listing will give the user a manual reference to his notification card file and would perhaps give him visual insights on how to improve his profile. The "front end" program of the 360 SDI system will edit various source documents into standard format and store them in a bulk storage device. We contemplate developing a retrospective retrieval system based around this bulk document file, but we are not setting a high priority on this project. We are convinced from user reactions and comments that the majority of scientists are not excessively concerned about retrospective retrieval as current awareness. Their biggest concern with the "scientific literature explosion" is to be aware of current research. Obviously scientists exploring new research areas would find retrospective retrieval useful; however, these people seem willing to tackle this searching manually if given a choice between current and retrospective retrieval. Ultimately, we hope to provide them with both capabilities.

The new system, programmed in COBOL (Common Business Oriented Language), will continue to serve as an experimental laboratory as has the present system. Additional statistical measurements are incorporated into this system for "more-in-depth" studies. We know profile significance values are a resource of intellectual effort, and we intend to derive methodologies to tap this knowledge bank for other information-related applications. Analysis of thesauri and keyword effectiveness and automatic retrospective retrieval query building are only two potential areas which can be developed while providing a valuable service simultaneously.

The relative percentage of interest computed as illustrated in Section IV of this article may tend to appear low compared with other percentages published about other SDI systems. However, our only criteria for measuring any degree of worthwhile service is through the reactions and comments of our users. Two specialized scientific listings (7, 8) are being compiled, one of these listings (8), through exclusive use of SDI, has accumulated 1,500 publications related to that specialized field in one year. It must be realized that this system does not replace existing methods or sources of literature searching, but rather supplements them by making more information accessible.

Acknowledgment

The author wishes to thank Mrs. Evelyn Delany of the Information Section, Computer Services, Ames Laboratory, for her programming efforts in plotting the graphs included in this article. To supply the data for these graphs, 83 supplementary graphs were plotted for each individual who participated during the 40-run production period.

References

1. SAGE, C. R., R. R. ANDERSON, and C. H. CAMERON, Ames Laboratory SDI Reference Manual, USAEC Rept. IS-940 (1964).
2. Experimental D.T.I.E. Magnetic Tape Service, Data Processing Office, Division of Technical Information Extension, USAEC, Oak Ridge, Tenn.
3. SAGE, C. R., R. R. ANDERSON, and D. R. FITZWATER, Adaptive Information Dissemination, *American Documentation*, 16 (No. 3):188, 190 (1965).
4. SAGE, C. R., and R. R. ANDERSON, *Ames Laboratory "Selective Dissemination of Information" Computer System—A Survey of User Evaluation.* To be submitted for publication.
5. SAGE, C. R., R. R. ANDERSON, and E. DELANY, *Ames Laboratory Selective Dissemination of Information General Information Preliminary Report,* Iowa State University, Ames, Iowa (1966).
6. MAPLE, C. G., Director, Computation Center, Iowa State University, Ames, Iowa. Private communication.
7. Rare Earth Information Center, Division of Technical Information, Ames, Iowa.
8. CAPPELLEN, J., H. J. SVEC, and C. R. SAGE, Bibliography of Mass Spectroscopy Literature, Compiled By Computer Method, USAEC Rept. IS-1335 (1966).

The Scholar and the Future of Microfilm

The reasons for the failure of roll film and microcards to be fully integrated into library practices are examined and compared with present practices in microfiche systems. It is concluded that microfilm systems in the recently introduced form of the fiche will finally be integrated into library usage.

R. R. DICKISON

*Oak Ridge National Laboratory**
Oak Ridge, Tennessee

In 1944 Mr. Fremont Rider (1) wrote a book about the exciting possibilities and potentialities of "micro-cards" as a solution for library growth problems. Rider noted (p. 92) that "We have had coming into our research libraries a mere trickle of micro-materials, where our micro-enthusiasts had hoped for, and had expected to have, a flood. And the reason why the flood has never come is the one just stated; that micro-reduction has never yet really integrated itself into library practice." Mr. Rider then went on to envision a micro-reduction system in which the catalog card for a book carried the micro-text of the book on its reverse side, mass-produced and inexpensive readers freely available to every user, and a circulation system in which a duplicate micro-copy of a book was given to a borrower, or, if he was willing to pay for it, a photostatic enlargement of the book. Mr. Rider argued persuasively that the new "micro-cards" could, if not within a few weeks then ultra-conservatively within two or three years, change the trickle into a flood.

In the next fifteen years from 1945 to 1960 certainly microcards received a fair trial in practically every library. One federal agency alone, the Atomic Energy Commission, made and distributed probably 20 million microcards before discontinuing their production altogether in July 1964. But, despite the considerable promotion and development, the dam never really broke. Just as roll film failed to integrate itself into library practice, microcards never came into general use in libraries.

In the past several years, another microform, the fiche, has made its appearance in American libraries. The question naturally arises, will the microfiche form integrate itself into library practice, or in fifteen years will it have failed just as roll film and microcards did?

One way to make a reasonable guess as to the future of microfiche is to look at the reasons given in the literature for the failure of roll film and microcard systems, and to see if these difficulties are still present in a microfiche system.

While a fairly wide variety of reasons is given for the previous failures, they can all be associated with one or the other of what appear to be the two major difficulties, lack of standardization and user resistance.

In addition to Rider, others, e.g., Scott (2), Tate (3), Riggs (4), and Piez (5), noted that lack of standardization was an effective barrier to the widespread use of roll film. Libraries received roll film perforated or unperforated, negative or positive, 16 mm, 35 mm, or 70 mm, and in a wide range of reduction ratios, usually anywhere from 8x to 30x. It is an understatement to say that this situation discouraged its use. Libraries were reluctant to purchase the wide variety of equipment needed to handle such diversity. Since libraries were reluctant to purchase, equipment manufacturers could not produce in quantity and prices remained high, which made libraries reluctant to buy.

User resistance was also an effective barrier to roll film. Users did not like being forced to go to the library for a reader and, once there, have difficulty in threading and using an unfamiliar machine. If these were overcome there sometimes remained the difficulty of locating a particular image on a reel containing hundreds or even thousands of images. If the desired image was located, usually an enlarged copy was desired, but there existed no convenient way to get a copy. In view of all the difficulties it seems a little surprising that roll film even survived in library usage. These difficulties have now, however, been overcome with the introduction of the cartridge reader and the several indexing devices associated with it for rapid frame location. Probably roll

* Operated by Union Carbide Corporation for the U. S. Atomic Energy Commission.

film in cartridges will make an impact on libraries in the next few years similar to the impact the microfiche form is now making.

Lack of standardization was apparently not as severe a problem with microcards, but still existed. Scott (2) noted of the micro-opaque system that it was "better than roll film for filing and retrieval and . . . closer to being a system than any other microform, but it is a system restricted to consultation of published material, impractical as a means of making a few copies only, difficult to reproduce and not sufficiently controlled by standards." There were at least three different sizes of microcards and at least five types of readers, none of which was inexpensive enough for libraries to make them freely available.

User resistance to microcards appears to have been about as formidable as it was to roll film. The quick and easy methods which Rider foresaw for duplicating microcards or making enlargements from them never materialized. More than one microcard user commented that all you could do with a microcard was read it, if you were lucky. The failure of microcards to be integrated into general library usage can probably be traced to the failure to devise good and/or cheap reader-printers.

Undoubtedly, the situation with regard to standardization is different with microfiche than it was with roll film or microcards. The efforts of the federal agencies to achieve standardization, begun in June 1963, have resulted in the first edition of "Federal Microfiche Standards," issued in September 1965 (8), to be used by all federal agencies and their contractors. These standards, which adhere to one of the two international standards for microfiche, appear to have had already a substantial impact and libraries are beginning to benefit from these efforts. The "elusive inexpensive portable film reader" (2, p. 490) is at last in sight. The ORNL library, and probably other libraries, has begun distribution in quantity of a simple desk-top fiche reader costing less than $100 per reader. More than 100 such readers have been distributed at ORNL. As has been noted (6), the cost, size, and simplicity of the reader is a key factor in a microfilm system.

Little has been published as yet about user resistance to microfiche. The survey (7) of 100 librarian users of the services of various federal agencies disclosed some dissatisfaction with microfiche, particularly with the quality. However, the experience of the ORNL library indicates that the objections most frequently raised to roll film and microcards do not exist in a microfiche system, and user resistance is consequently disappearing. Several of the information centers have converted their entire files to microfiche and one is supplying duplicate microfiche copies of documents to its clientele rather than bibliographic references to the documents, except, of course, for copyrighted material. A microfiche can be easily and cheaply duplicated and currently about 450 a week are being duplicated and distributed to microfiche users for their files. An automatic step and repeat enlarger furnishes hard copy quickly and cheaply if it is required. Users have reported no difficulties in placing the film in the reader, in locating the desired image, and in obtaining enlargements if needed. About 20 conveniently located reader-printers which provide quick enlargements have been distributed.

Not all of the difficulties with microfilm have been resolved with the microfiche. "Microfilms are microfilms and not the original book" (9), and this difficulty appears incapable of resolution. Not all of the necessary equipment for a complete microfiche system is commercially available, but it is reasonable to expect that it will appear and that the costs of the entire system will go down. The quality can and probably will be improved.

It seems not unreasonable to predict now that microfilm, in the form of the fiche, will finally be integrated into library practice and the exciting possibilities for solving library growth problems, which Mr. Rider envisioned over 20 years ago, are now at hand.

References

1. Rider, F., *The Scholar and the Future of the Research Library*, Hadham Press, New York, 1944.
2. Scott, P., Advances and Goals in Microphotography, *Library Trends*, 8:458–492 (1960).
3. Tate, V. D., An Appraisal of Microfilm, *American Documentation*, 1:98 (1950).
4. Riggs, J. A., The State of Microtext Publications, *Library Trends*, 8:379 (1960).
5. Piez, G. T., Microfiche Standard Adopted, *Special Libraries*, 55:390 (1964).
6. Gray, D. E., Practical Experience in Microfacsimile Publications, *American Documentation*, 3:58–61 (1952).
7. SLA Government Information Services Committee, Users Look at Information Centers, *Special Libraries*, 57: 45–50 (1966).
8. Federal Microfiche Standards, P.B. 16730. 1st Ed. September 1965.
9. Jackson, W. A., Some Limitations of Microfilm, *Papers of the Bibliographical Society of America*, 35:281–288 (Fourth Quarter, 1941).

PICS: The Pharmaceutical Information Control System of Merck Sharp and Dohme Research Laboratories

The Pharmaceutical Information Control System (PICS), developed at Merck Sharp & Dohme Research Laboratories, provides centralized control and methodology for a series of decentralized information areas in the Division. It is compatible with and instrumental in total data processing and analysis of research information. Serving as a Core Index to all information resources of the Research Laboratories, it also processes, stores, and retrieves research project information for the staff members for planning and retrospective searches. A register of all domestic and international clinical research information on experimental and in-line products of Merck & Co., Inc., is provided by the system.

An eight-digit dual-faceted classification code was developed based on a companywide program identification scheme. This code of two mutually exclusive facets enables us to identify a product with a field of research. This code has been adopted for use in administrative planning, cost accounting, time allocation, and internal reporting. The uniform use of this information code within the Research Division minimizes the vocabulary barrier between the user and information system and provides the system with a self-indexing device for internal reports.

Incoming mail is copied and registered by the information center prior to transmittal to the addressee. Copies of all outgoing mail and intramural correspondence are directed to the center. An information scientist analyzes each document and selects the project code and document descriptors. This information is punched into 80-column cards. The documents are filed by code and the cards are filed alphabetically by name, term, and by date. Output forms include information, documents, printouts of document citations, and reports to drug regulatory agencies. Continuous system evaluation results in reduced I/O time, better utilization of personnel, and improved user feedback and contact.

MARGARET C. KOLB, JEROME T. MADDOCK,
and BARBARA N. WEAVER

Merck Sharp & Dohme Research Laboratories
Rahway, N. J.

• Introduction

One of the major problems facing industrial and governmental organizations at present is the mushroom-like growth of many totally incompatible "mini-systems" for handling particular aspects of the total information picture of that organization. Each of these "mini-systems" fulfills its responsibility to the organizational unit it serves directly, but fails to fulfill its responsibility as an integral part of the total information system of the organization. The need for compatability and integration of systems is evident, and must be a primary consideration of good systems design work. This paper describes a system which was planned to meet these requirements, as well as the modular growth requirements of an industrial organization.

The Pharmaceutical Information Control System (PICS) has been developed in the Merck Sharp & Dohme Research Laboratories to provide centralized control and methodology for the series of decentralized information areas in Rahway, New Jersey, and West Point, Pennsylvania. It provides a significant quantity of relevant intramural information to various levels of managerial and operational employees without jeopardizing the security of the company's informational resources. The document control established by this system serves as an information bridge to the complex Division-wide activities of project planning and data processing. Uniform use of a classified code, based on a program identification scheme, permits the accumulation of a total Project Profile by relating and correlating costs, manpower, labora-

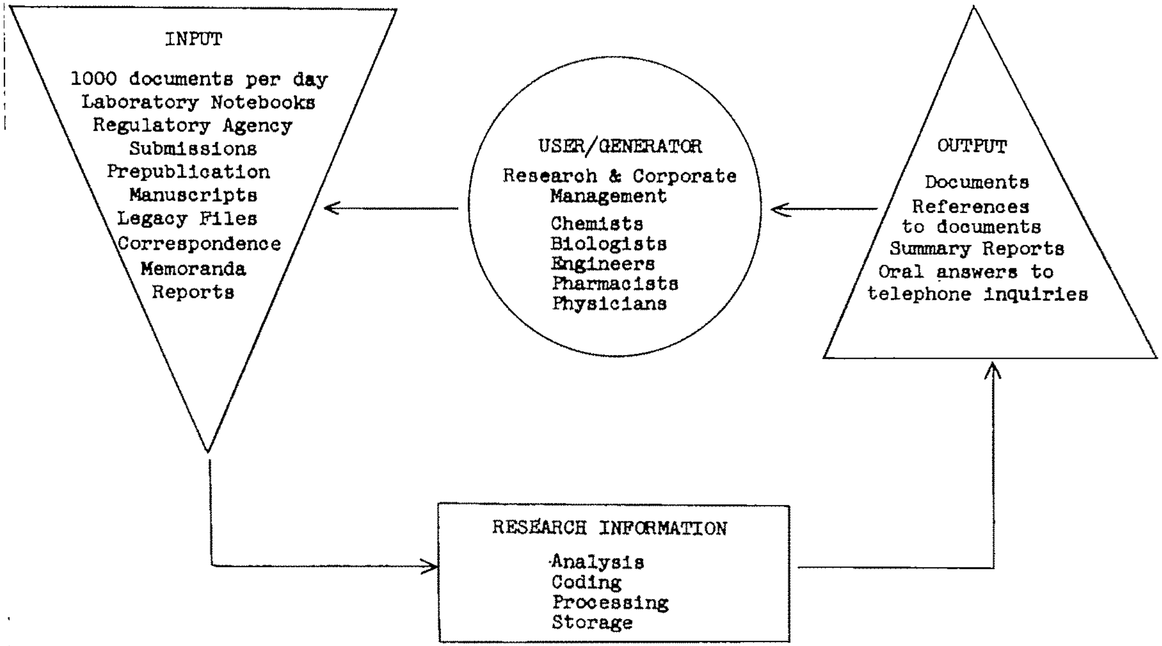


FIG 1. PICS System

tory results, and information generated during the life of a research project. Thus, information has become a measurable commodity which can be evaluated along with other products of research. The system produces a Core Index to the Information Resources of Research which includes not only resources within the Information Centers, but also the resources of staff scientists through the incorporation in the system of an index to individual specialized collections. PICS is organized on a staff level within the Division. It provides consultant services to the user groups to help solve their individual information network problems.

The user group consists of chemists, biologists, engineers, pharmacists, physicians, and research and corporate management. The system input is approximately 1,000 documents per day, composed of reports, memoranda, correspondence, laboratory notebooks, regulatory agency submissions, prepublication manuscripts, and legacy files from various directors and departments. Published information is excluded from this system; it is processed elsewhere in the Division. Output may be either in the form of a summary report prepared by an information scientist, documents, references to documents, or oral answers to direct questions by telephone. The choice of low cost, simple EDP equipment for PICS was purposeful to maintain economy as well as efficiency. PICS became operational in April 1963; a description of the system in use earlier at Merck is contained in *Information and Communication Practices in Industry* (1).

• Classification Code

An eight-digit, dual-faceted classification scheme has been designed for Division-wide project identification. Management utilizes the code for cost accounting, time allocation, and internal reporting. The Information Center uses the code for information storage and retrieval. The first four-digit facet defines a research program; the second four-digit facet, a serial code, identifies a product. This code of two mutually exclusive facets relates a product with its field of research. Tag facets are used as needed to identify combinations, formulations, and routes of administration of formulations. A numeric subdivision, when added to this project number, further specifies categories of information and permits the retrieval of chemical, biological, marketing, or other specific aspects of a project or product.

• Vocabulary Control

The West Point and Rahway Research Information Centers are evaluating controlled vs. uncontrolled vocabulary.

The keywords from text approach is employed in Research Information-West Point to test the hypothesis that within this controlled homogenous scientific community of Clinical Research, the level of terminology is relatively constant and is self-controlled because the users are also the generators of the documents. This approach is being critically tested for growth of vocabulary, efficiency of retrieval, and economics of input against Research Information-Rahway. There, vocabulary is rigidly controlled by an intricate term and contact thesaurus-based system which serves a larger, more heterogeneous scientific community.

• Document Processing

Incoming mail is opened and copied on a Xerox 914 for the Information Center before transmittal to the addressee. Carbon copies of all outgoing mail and intracompany reports and memos are directed to the Center. Thus, a copy of every research document, internally or externally generated, is processed and stored in the Research Information Center, ensuring strict input control over the Division's information resources.

All documents are analyzed and coded by an Information Scientist, who writes the code directly on the document and marks any appropriate additional descriptors directly on the text (with a nonreproducible chromatic pencil). For large sets of identically coded documents, such as form letters, the slave typewriter unit of the 870 system is programmed "on" at the code field to generate the document code on labels simultaneously with punching of the registration card. This same feature of the 870 is used to automatically print all labels for the more than 5,000 folders added to the file each year. Recently, case report forms have been preprinted with the project code number to reduce processing time. Complete document description is punched into 80-column IBM cards with field definitions for contacts (individual or organization names) or terms, dates of documents, initials of addressees, authors or correspondents, and project code.

For every document a date registry card is punched, plus any necessary term and contact cards. These cards are identical, differing only in the first field (columns 1–25) when necessary.

The chronological "date card deck" acts as a complete document registry of the Information Center. Two additional auxiliary alphabetic decks are maintained as indexes to the individual terms and contacts for retrieval.

1. The term deck contains all document descriptors, entries in the project code thesaurus, cross references to preferred synonyms and preferred abbreviations, and the terms from auxiliary collections.
2. The contact deck contains cross references to preferred individual or organizational identification, preferred abbreviations, names of individuals affiliated with organizational contacts, and names derived from the document.

All punched cards are sight-verified against documents. The verifier marks the document to indicate that it has been completely processed and marks the cards to indicate the index into which they will be filed.

• Document Registration of Clinical Statements of Investigators and Case Reports

Completed Statements of Investigators (U. S. Food and Drug Administration Forms FD 1572 and FD 1573) are received in the Information Center prior to transmittal to the Staff Clinicians. Each Statement of Investigator is coded, indexed, and assigned a serial registry number which then becomes that investigator's identification number for that specific clinical study. This number is used to identify all case reports submitted by this investigator in reporting results of his study, and is carried through in the computer evaluation of the study results. A complete Statement of Investigator description, including principal Investigator's name, date of Statement, registry number, initials of the responsible Merck clinician, and product code, is punched into IBM cards in triplicate, with additional cards for any descriptor terms, and all other investigators participating in the study. The format of these cards is consistent with the format of those cards previously described. Cards are filed by investigator in the contact index and by product code and registry number in the Statement registry. Upon approval and duplication for IND inclusion, the S.I. form is returned to Research Information where the registry card is signal-punched to indicate that the form has been processed and that copies have been sent to F.D.A. The original S.I. form is filed by investigator, and a copy is filed by product code with that investigator's case reports.

All case reports are registered by Research Information upon receipt from the investigator. A serial registry of these reports is maintained on cards for each investigational drug. Each case report form is stamped with the investigator's S.I. number and the serial patient number. A copy of the registered case report is forwarded to the Merck Clinician. The original is retained in the Information Center. From this original, one card is punched containing the investigator's S.I. number, patient number, patient name, age, sex, date of report, initials of the Merck Clinician, and product code. The patient name is restricted to four characters, utilizing the method developed by Dr. John Tukey for abbreviations. On controlled studies (i.e., double blind, crossover, etc.), patient numbers from the protocol for the study are included as well. Manuscripts, letters, etc., which contain clinical commentary are also registered as case reports. A single alphabetic character punched into the card indicates the report type. The cards and documents are filed by product code.

Up-to-date case report registries and Statement of Investigator registries on any product are produced on demand for the clinical staff and for preparation of reports to government regulatory agencies through the use of the 870 Document Writing System.

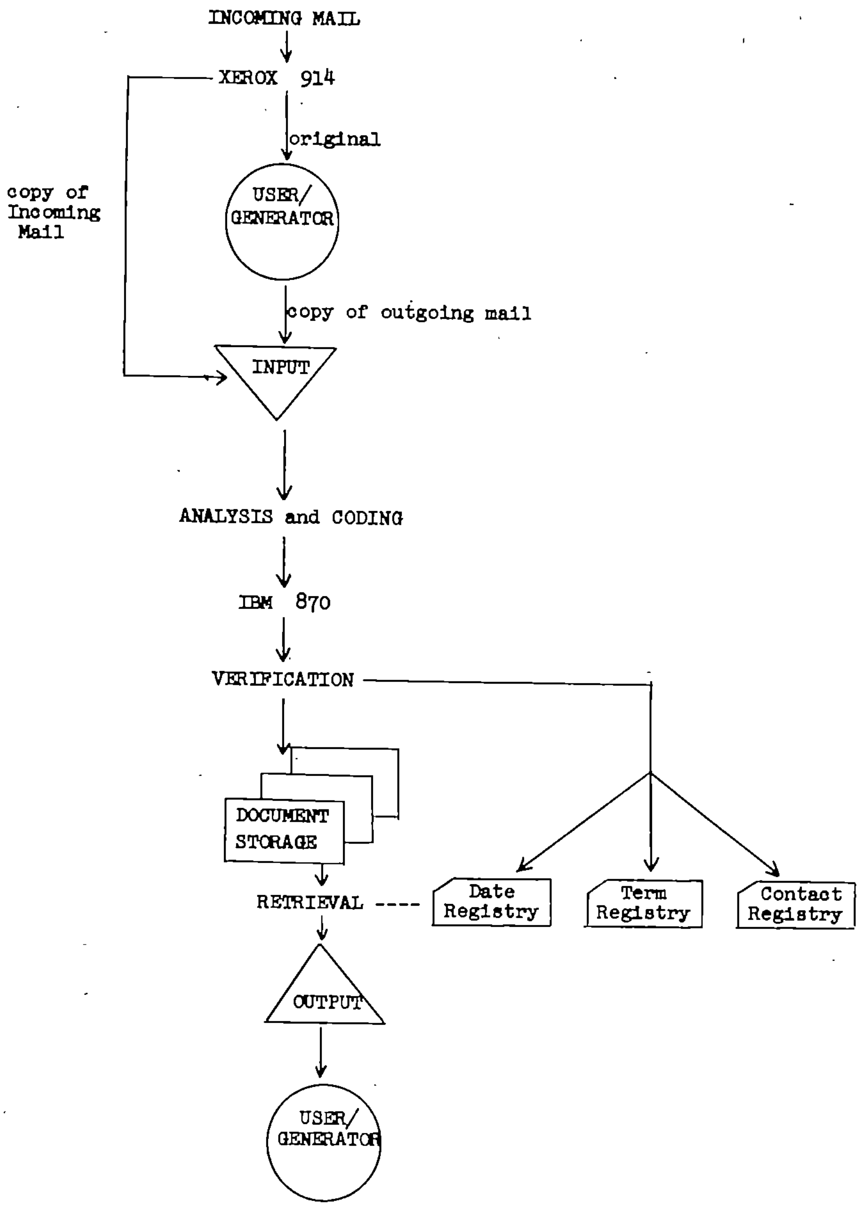


FIG. 2. Document processing

- **Storage**

Documents are stored in macroform chronologically by project and code subdivision.

Legacy files from various directors and departments are stored in microform after they have been entered into the Core Inventory of Total Information Resources of Research, which is maintained in the form of an IBM card deck. This microform storage eliminates the unnecessary build-up of files within the Division, yet maintains their personal retrievability by filming them in essentially the form maintained by the individual or department.

Laboratory notebooks are issued by the Research Information Center. Completed notebooks are stored in the Center, in macroform, with a microform copy stored at another location for security.

- **Retrieval**

Search requests are normally initiated through a telephone call or a personal visit from the user to the Center. At the present time, approximately 50 searches per day are processed by information scientists. Searches are done manually or mechanically, depending upon the nature of the request. Information, citations, and/or documents are retrieved through the Core Index by any

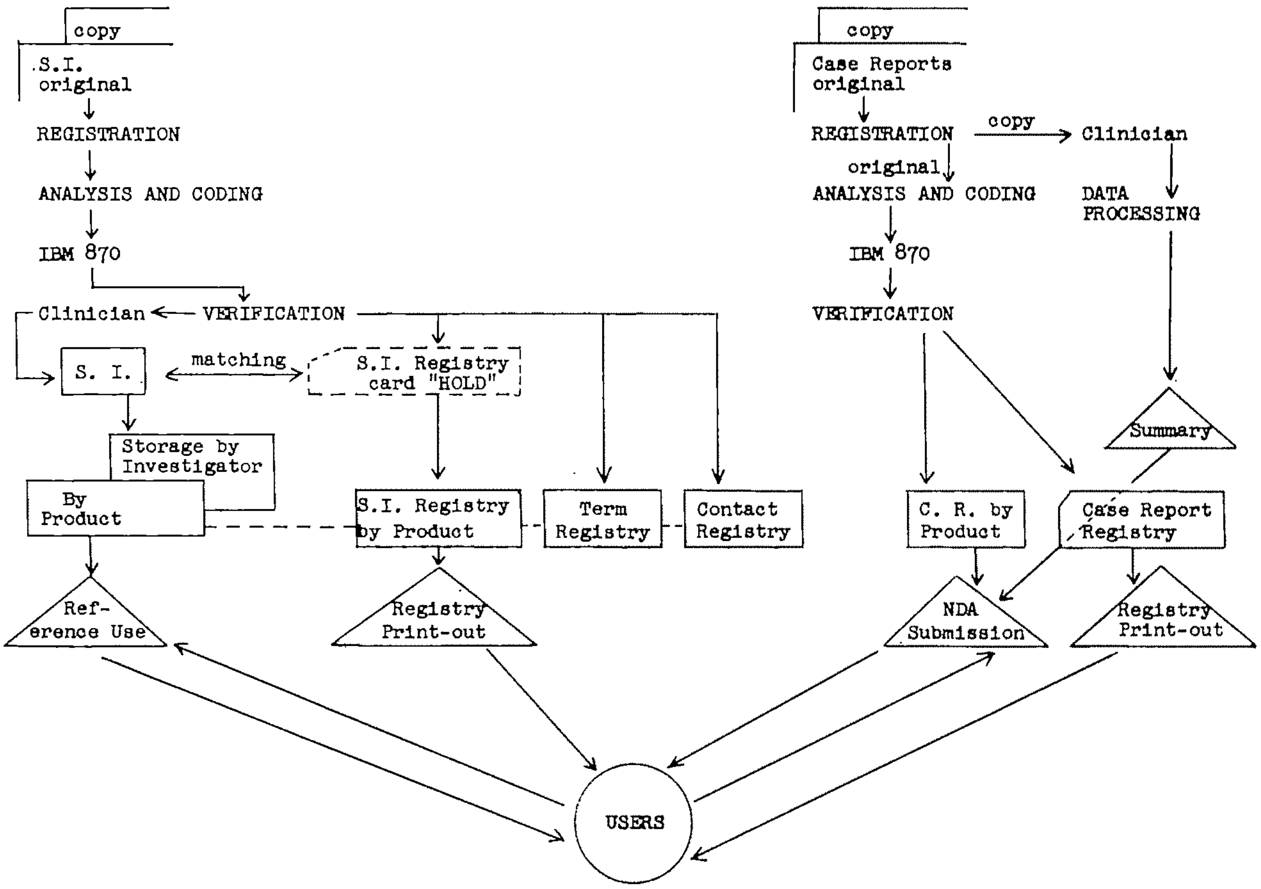


FIG. 3. Document processing of statements of investigator and case reports

one or combination of the document description aspects.

The questions range from the correct spelling of a name or chemical compound to the evaluation of past research results.

One of the most valuable services of PICS to Merck management is the extremely rapid output of retrospective searches on an entire research project or a specific aspect of a project from the resources of its information bank. Since the documents are filed in project-classified arrangement, they are pre-ordered for this use with zero time delay.

Concept searches are often initiated by scientists pursuing a new approach to a laboratory problem. As one of the possible search terms is traced from Core Index to its location in file, other relevant documents utilizing vocabulary more generic or specific are automatically retrieved due to the classified-document arrangement. Thus, the project arrangement complements the Core Index as a retrieval device.

Another important output form generated by Research Information at West Point is the master set of case reports for the clinical portion of New Drug Applications (NDA's) and periodic reports to the U. S. Food and Drug Administration. The case report registry is used to verify the completeness of the case reports in the product file and the processed clinical data. The information scientist meets with the responsible staff clinician to discuss the proposed format, arrangement, and inclusion of supporting documents other than case reports, and to obtain final approval for the master copy before it is released for duplication.

The Registries now produced on the IBM 870 will be generated from magnetic tape within a short time.

• Summary

PICS is compatible with and instrumental in total data processing and analysis of research information. Serving as a Core Inventory to all information resources of the Research Laboratories, it provides research project information to staff members for planning and retrospective searches. The uniform use of the basic eight-digit code throughout the Division furnishes a self-indexing device for internal reports and minimizes the vocabulary barrier between the user and the system. A register of all domestic and international clinical research information on experimental and in-line products of Merck & Co., Inc., is generated by this system.

Unique operational features of PICS include the absence of any work sheets for processing documents, the addition of subsequent source references to any document cited to provide effective bibliographic coupling, the automation of many clerical operations through the use of the IBM 870, and the modular system design which offers both manual and machine retrieval capabilities and allows for the transition to more sophisticated equipment. Microform input and the expansion of the Core Index are contemplated for the future.

Within PICS, functions are constantly undergoing evaluation. Techniques, such as user interviews, retrieval-time counts, document-input counts, and time-lag averages using punched-card methods, are being used for reduction of I/0 time, better utilization of personnel, and improved user feedback and communication. The results of these studies will be reported in a future paper.

References

1. Kolb, M. C., Chemical Research File Departments as Information Services, *in* T. E. R. Singer, ed., *Information and Communication Practice in Industry,* Reinhold, New York, pp. 118–138, 1958.

A Computer-Program System to Facilitate the Study of Technical Documents[1]

Symbiont is a computer-and-program system for use in research on computer-aided study. It stores, retrieves, and displays documents and parts of documents. It "semi-automates" the taking of verbatim notes. It facilitates the manipulation and intercomparison of graphs. And it conducts searches for passages of text that contain specified words or phrases. Experience with Symbiont and plans for its improvement are described.

DANIEL G. BOBROW,[2] R. Y. KAIN,[3]
BERTRAM RAPHAEL,[4] and J. C. R. LICKLIDER [5]

• Introduction

The purpose of this paper is to describe a system, consisting of a digital computer [6] and a computer program, intended for exploration of man-machine interaction and computer assistance to man in the study of technical documents.

The system provides a physical study situation that includes a desk, an electric typewriter,[7] a display screen, and a light-sensitive pointer or stylus ("light pen").[8] The user of the system, whom we shall call "the student," requests services and controls operations by typing command characters or symbols on the typewriter or by touching illuminated areas of the display screen with the light pen. The computer and program system, which we call "Symbiont" because we hope to develop it into a truly symbiotic partner of the student, displays information to the student via the typewriter or the display screen. The display screen, which is a 10-inch square area on the face of a cathode-ray tube, represents alphanumeric symbols and graphs. Whenever part of a displayed pattern is touched by the tip of the light pen, the computer can tell what part was touched and when. The combination of computer-controlled cathode-ray display and computer-signaling light pen is a convenient and flexible arrangement for man-computer communication.

Symbiont is an early stage of what we hope will be a continuing evolution. However, a sufficient set of functions has been implemented to lead us to take stock and gain experience in their use before modifying existing functions or adding new ones.

Inasmuch as Symbiont is an exploratory tool, for use mainly by students who are at the same time experimenters, we have not considered it necessary to perfect or polish. For example, the display flickers. With the aid of character-generation and display-buffering equipment, we could achieve a steady display: the technology is far enough advanced to fulfill the display function well. At present, however, the equipment required for flicker-free display is expensive, and we prefer to put available funds into other things. Our reasoning is that, in due course, good, steady displays will become relatively inexpensive, and in the interim we can make allowances for a bit of flicker. The same argument applies to text storage capacity, text searching rate, and production of permanent copy. In short, our aim has been to realize several interesting functions now, even though in ways for which certain allowances have to be made, in order to gain early experience in using the functions and to provide a basis for practical system design when advances in technology make it possible to implement the functions effectively and economically.

[1] The work reported here was supported by the Council on Library Resources at Bolt Beranek and Newman, Inc.

[2] Bolt Beranek and Newman, Inc.

[3] Assistant Professor of Electrical Engineering at M.I.T. and Consultant to Bolt Beranek and Newman, Inc.

[4] At present, at Stanford Research Institute; formerly Consultant to Bolt Beranek and Newman, Inc.

[5] At present, at the IBM Research Center; formerly at Bolt Beranek and Newman, Inc.

[6] Programmed Data Processor 1 (PDP-1), manufactured by the Digital Equipment Corporation, Maynard, Massachusetts.

[7] IBM Executive Electric Typewriter (Model B), modified for use with digital computer by the Soroban Engineering, Inc., Melbourne, Florida.

[8] The display screen and light pen are parts of the PDP-1 system, supplied by the Digital Equipment Corporation.

• Operations and Functions Implemented

A study session with Symbiont starts with the computer turned on, the basic program running, and the text and graphs of several technical documents already punched into machine-readable paper tape. The text is represented character-by-character in a standard alphanumeric code. The curves of the graphs are represented numerically by coordinate values at selected points along the abscissae, and the calibrations, labels, and legend are represented alphanumerically in a prescribed format.

At the beginning of his study session, the student loads representations of the documents he plans to study from an input tape into the computer memory. Then, typically, he calls for a document and reads or scans it. He calls for it by typing any part of its standard bibliographic citation that specifies it uniquely—the author's name, for example, or a major part of the title, or the name (and perhaps volume or year) of the journal in which the document was published. Symbiont finds the specified document and presents the first screen-page of it. (A screen-page is about 150 words in length. Lines and pages have to be shorter on presently available display screen than full lines and pages are in most document-pages.) The student turns pages in the forward direction by hitting the space bar of the typewriter. He may back up a page at a time by hitting the backspace key.

While reading or scanning, the student comes upon a passage that he wants to record verbatim for future reference—a passage he would ordinarily copy onto a note card. With the aid of Symbiont, he records it on paper tape or in the note-file part of the computer memory. To punch it on paper tape, he touches the initial printed character or characters of the passage with the light pen and then types "b" (for "begin"). Underlining thereupon appears beneath the character(s) touched. Then he touches the final printed character(s) of the passage and types "e" (for "end"). Underlining thereupon appears beneath the ending of the passage, and immediately spreads back to the beginning. The passage is thus singled out for inspection by the student and for action by the computer. When the student types "p" (for "punch"), Symbiont punches the passage into paper tape. If the student next underlines the bibliographic-citation string that appears at the head of the document, Symbiont appends the citation to the note, thus handling a chore that ordinarily plagues the conscientious notetaker when he takes his notes and the unconscientious notetaker when he tries to use his notes. The student can string any number of passages together by underlining them and punching them one at a time, in groups, or all at once.

If the student prefers to note the passage in the computer memory instead of paper tape, he needs to specify a "tag" with which to retrieve it. He specifies the tag (before underlining the passage) by typing "t" (for "tag") and then any symbol, or indeed any string of printing characters and spaces, terminated by a carriage return. He then underlines the passage and types "n" (for "note"). Alternatively, he can assign to the passage a "label," which is functionally equivalent to a tag, but specified initially by underlining a string of characters on the screen with the light pen and then typing "l" (for "label"). The procedure for connecting the label to its passage is the same as the procedure for connecting a tag. Tags and labels go into a "glossary" of retrieval terms associated with the note file. To see what the glossary holds at any time, the student types "g" and looks at the screen. If the glossary is more than one page long, he turns its pages as though it were text.

Often the student wants to retrieve notes, and sometimes he wants to amend or combine them. To retrieve a note, the student types "r" (for "retrieve") and then types the tag or label (or if more convenient, designates a corresponding string of characters by underlining them with the light pen). In amending and combining retrieved notes, the student is constrained by the present system to serial designation and concatenation of passages and subpassages. Under these constraints, editing is like operating a switch engine. However, it will be easy to introduce the operations of deletion and insertion.

Verbatim notetaking and retrieval of notes are admittedly minor matters. More vital is retrieval of primary information. In the present context, since the student is assumed to be working with a small collection of documents known to be relevant to the topic under investigation, the retrieval problem is not primarily one of finding documents. It is primarily one of finding passages in documents that discuss particular ideas, passages that are relevant to particular technical points. The approach of Symbiont to this problem is to automate the scanning of text for specified configurations of retrieval terms.

Symbiont carries out searches with reference to one, two, or three sets of retrieval terms. Each set may contain any number of terms of any length. For retrieval purposes, all the members of a set are assumed to be synonymous: Symbiont considers that it has found the set as soon as it finds any member of a set. Symbiont looks for members of the three sets within a "neighborhood" of text. A neighborhood is n lines in length, and the student can set n to any value he likes. Five lines make a good neighborhood.

Before conducting a search, the student types "t" (for "terms"), then types the strings of characters that constitute the alternate terms of the first retrieval set, and types "1" to designate this set as the first. Then the student types "t," the terms of the second set, and "2," and finally "t," the terms of the third set, and "3." The three sets might be, for example:

1	2	3
cigarette	lung	cancer
cigarettes	lungs	carcinoma
cigar	pulmonary	
cigars		
pipe		
pipes		
tobacco		
tobaccos		
nicotine		

The student then decides whether he wants a passage (neighborhood) dealing with one of the three, two of the three, or all three ideas (sets), and he initiates the search by typing "f1," "f2," or "f3" (for "find one," etc.). Symbiont thereupon searches serially through the text until it either comes to the end or finds a neighborhood that meets the specifications. If it comes to the end, it displays "not found." If it finds a neighborhood that meets the specifications, it displays on the screen the text containing the neighborhood, showing a small amount of preceding text and a larger amount of succeeding text. The student may turn pages, copy passages, etc., in the way described earlier, or he may type "f1," "f2," or "f3" and have Symbiont look for another passage that also meets the specification.

Although the idea-retrieval technique just described is primitive, it is surprisingly effective if the student is clever in setting up the sets of terms. Typically, the student starts with a loose retrieval prescription and tightens it as he makes his way through his collection of documents.

Graphs are composed by the computer from tabulated data and presented on the screen as graphs. They are displayed separately from text. They have keys that associate labels with curves; they have calibrated and labeled axes; and they have legends. Curves are approximated by straight-line segments, dashed and/or dotted in eight patterns. A family of curves can have any number of members, but in the present system, only one label. Up to eight families of curves can be superimposed upon one grid. Two grids can be set side-by-side to facilitate comparison. If the graphs are fundamentally comparable but different in scale factor, the student can, with the aid of the light pen, expand or compress the scales of one or the other until the two presentations are directly comparable. He adjusts the length or position of a line segment of the coordinate frame by touching one of its ends with the light pen (which "picks up" the end-point) and then moving the end-point to the desired location. If necessary, he repeats the procedure with the other end-point. The computer then rescales and relocates the entire graph. If two graphs are displayed side-by-side, one of them can be moved and superimposed upon the other, or curves can be transferred from one to another. These operations facilitate synthesis of a composite picture from results obtained by diverse investigators.

Symbiont makes it easy to modify not only the size of a graph but also the grid structure, the structure of the subdivision of the area within the graph. When it changes a grid, it also changes the numbers associated with the grid lines (i.e., the numbers associated with the scale-calibration points).

At the bottom of the screen, there is a display of numerals and control symbols. By pointing with the light pen to individual numerals in proper sequence, the student can build up any number he needs. Then, designating with the light pen the control symbol "SCALE" and a scale-calibration point he can substitute the assembled number for the number theretofore associated with the scale-calibration point. As soon as new numbers have been associated with two calibration points on a linear axis scale, the computer substitutes new values at all the other calibration points on the axis.

If he wants to change the number of grid lines that subdivide (say) the "pressure" scale of a graph, the student points with the light pen to the control symbol "GRID" and then to the label "PRESSURE" and then to the appropriate numeral corresponding to the desired number of grid lines. The computer immediately redraws the grid, leaving the extreme grid lines unchanged, and substitutes the appropriate new numbers near the intersections of the new grid lines and the horizontal axis. With these procedures, the student may experiment rapidly with various frames and grids, for he need specify only the essential parameters of each coordinate system. As soon as they are specified, Symbiont develops the detailed pattern.

• Evaluations and Plans for Improvement

Our experience in using Symbiont has been limited by shortage of input tapes and by smallness of the computer memory. A semi-automatic tape-preparation subsystem and an arrangement for moving information automatically between primary (core) and secondary (drum) memory are the items of highest priority in the plans for Symbiont II. Even on the basis of the limited experience, however, it seems clear to us that the functions provided by Symbiont I (the system thus far implemented) are effective as aids in technical study. The function of searching for ideas, as primitive as the implementation is in Symbiont I, is little short of powerful. The automation of verbatim notetaking, despite shortcomings in human engineering, seems capable of serving as the foundation for efficient personal documentation systems.

In Symbiont I, however, too many of the graph-handling functions deal with frames, grids, and labels, and not enough deal with curves. The limitation to linear transformations is highly constraining. We must admit, therefore, that the graph-handling functions of Symbiont I do little more than (a) afford convenience in the few parts of the over-all process of graph manipulation that they subsume and (b) make it seem plausible that a fuller set of functions (involving perhaps 10 times as much programming) would be truly useful.

The plans for Symbiont II call for the following modifications of, and additions to, Symbiont I:

1. A subsystem to "semi-automate" preparation of input tapes of textual and graphical information. Because performance of the system during study does not depend upon how the tapes were prepared, we deferred work of a tape-preparation subsystem and relied upon manual production of input tapes. Manual production proved not to be satisfactory. For Symbiont II, we plan to take text mainly from monotype and linotype tapes and to use computer film-reading techniques in converting graphical data to tabular form.
2. Extension of the storage areas, confined to core memory and supplementary paper tape in Symbiont I, to the magnetic drum (22 times 4,096 18-bit words) now associated with the PDP-1, and perhaps also from the drum to magnetic tape units.
3. Substitution of light-pen for typewriter control of most operations that deal with information displayed on the screen.
4. A descriptor-and-thesaurus system for retrieving documents from store. Symbiont I retrieves documents with the same searching system it uses in finding passages. (A bibliographic designation precedes each document in the store of text.) That will be too slow when the store becomes large.
5. A scheme for turning several or many pages at a time or for going immediately to a particular page specified by page number.
6. More reliance upon predetermined sequences of manipulation and less upon control characters. For example, to underline a segment of text, it should suffice to point with the light pen to an "underline" light button, then to the beginning of the passage, and then to the end of the passage. It is an unnecessary nuisance to have to specify "end" after having specified "begin." However, streamlining the procedure in this way will make it necessary to provide a way of reminding the student when he forgets where he is, in a sequence of operations, and a way of letting him linger on (or return to) a particular operation long enough to correct a mistake in specifying it.
7. Handling of notes precisely as though they were documents. Notes will be permitted to contain graphs. The note-retrieval glossary will be associated with the document-retrieval system.
8. Acceptance of notes phrased by student. This now seems essential even though it is easy for him to record verbatim notes.
9. Provision for extraction from text of individual words, individual phrases (delimited by punctuation marks), individual sentences, and individual paragraphs merely by pointing. It is an unnecessary nuisance to underline (i.e., to point to both ends of) a segment unless one wants to extract a sequence of characters that does not constitute a formal unit.
10. Labeling of individual curves as well as of families.
11. Labeling near the curve as an alternative to associating label and curve by key.
12. Search for more than three sets of terms, and for other combinations (such as 1 and 2 or 1 and 3) than any m of n.
13. Storage and retrieval of the sets of terms used in searching text. It is not good to have to type a set of terms more than once, and it will be easy to store them for future reference. The student will be able to retrieve a set by typing any term in the set. Symbiont II will display all the sets that contain the typed term and let the student select the one he wants by pointing to it.
14. In designating parts of graphs to the program for action, more pointing to the parts themselves, and less pointing to their names.
15. Transformation between linear and logarithmic coordinates.
16. Fitting of curves (specified by type, such as sine, exponential, and power series) to tabulated numerical data, and determination of goodness of fit.
17. Weighted averaging of curves.

The present plan is to effect the foregoing improvements, to gain further experience, and then, in proceeding to the third generation of study facilities, to meld them with arrangements, not described, to facilitate the organization and retrieval of notes and data and the preparation of technical papers. For further information about the context of the Symbiont system, see reference (1) below.

Reference

1. Licklider, J. C. R., *Libraries of the Future*, M.I.T. Press, Cambridge, Massachusetts, 1965.

Biological Dictionary Preparation, Control, and Maintenance

A description of the working processes involved in preparation, control, and maintenance of a biological dictionary or thesaurus is given. The actual use of the dictionary by the abstractor-indexer is illustrated by sample coding sheets and examples from the dictionary of cross references, instructions, and scope notes. The use of specific authorities such as the World Health Organization Classification of Diseases is cited. Emphasis is placed on firm control by one person trained in biological science, especially in the matter of synonyms and of family or generic entries. Without this control, a great many pertinent references can easily be lost when a search is made. Non-thesaurus, uncontrolled indexing is very wasteful and unreliable.

S. JANE WEINSTEIN

Abbott Laboratories
North Chicago, Illinois

Control of indexing terms is absolutely essential when a group of subject-trained individuals is responsible for both storing and retrieving information.

The language used by the originator of the article being coded may not and often does not match that used by the indexer or the searcher (1). Communication and language problems involve viewpoint or class context, generics, and semantics (1).

Since our indexers work with one viewpoint in mind (the effects of drugs on biological systems), this factor is not so important as the other two, semantics and generics.

The relationship between words and their meanings such as synonyms, near synonyms, and homographs make up the semantic problem (2).

An information system must show how words are used (3). It must also make provision for cross reference to enable the searcher to retrieve all pertinent information on concepts of interest to him (4).

The need for and concept of a technical thesaurus has been demonstrated by others (5–8). Thesauri and/or authority lists have been successfully created by other organizations (9–16), but none of these were specific enough or broad enough for the needs of our group.

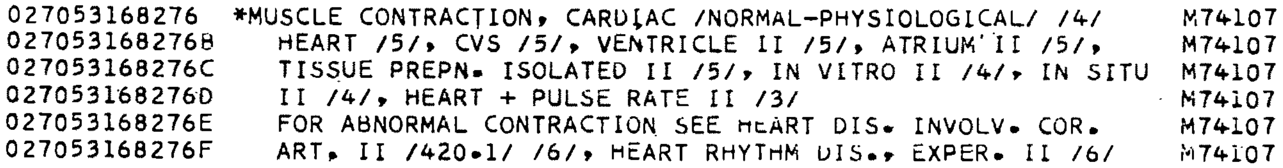
Early compilations or authority lists were often called dictionaries rather than thesauri. The Abbott Abstracts biological authority list still bears the label of dictionary although it is fulfilling the function of a thesaurus by the inclusion of scope notes and cross references as needed.

Our methods of preparing a dictionary-thesaurus have proven to be reliable, uniform, accurate, and extremely efficient (which means that they are excellent time savers and as such save money).

The Abbott Abstracts Biological Dictionary was organized in 1960 to control the indexing terms used by the group of subject specialists who abstract and index the current published literature for the working scientists at Abbott Laboratories (18). Our group of subject specialists, who number seven at the present time, include two Ph.D.'s, one a biochemist, the other an organic chemist. The other members of the group all have master's degrees or the equivalent in one of the sciences. They include a physiologist, a pharmacologist, a zoologist, and two other organic chemists.

The Biological Dictionary lists in alphabetical order terms, cross references, and all related or generic terms required when a specific concept is used, and provides scope notes whenever necessary to clarify the intended meaning of a term.

An illustration of a term can be seen in Fig. 1. The term "Muscle Contraction, Cardiac/Normal-Physiological" requires the terms "Heart" and "CVS" (cardiovascular system) to be indexed as well. The generic relationship is thus established and maintained. Optional terms that may be used by the indexer if they are discussed in the article are appended as scope notes. The symbol "II" after each of these terms means "if indicated" to stress that the listed terms are optional and



```
027053168276   *MUSCLE CONTRACTION, CARDIAC /NORMAL-PHYSIOLOGICAL/ /4/      M74107
027053168276B    HEART /5/, CVS /5/, VENTRICLE II /5/, ATRIUM II /5/,       M74107
027053168276C    TISSUE PREPN. ISOLATED II /5/, IN VITRO II /4/, IN SITU    M74107
027053168276D    II /4/, HEART + PULSE RATE II /3/                          M74107
027053168276E    FOR ABNORMAL CONTRACTION SEE HEART DIS. INVOLV. COR.       M74107
027053168276F    ART, II /420.1/ /6/, HEART RHYTHM DIS., EXPER. II /6/      M74107
```

FIG. 1. Dictionary term, generic entry, "Muscle Contraction, Cardiac /Normal-Physiological/"

not obligatory. A cross reference is also made to "Abnormal Contraction."

Synonyms are carefully controlled. The scope of this dictionary-thesaurus is broad in that it includes all concepts needed for indexing or coding articles written about the effects of drugs on biological systems, with the exception of the drugs themselves and the generic chemical classes related to those drugs.

The classified concepts (the most frequently used terms of which are printed on coding sheets for the indexer's use) are divided into 10 categories. (Fig. 2.)

The 10 categories include two groups of terms that are controlled in a chemical dictionary by an organic chemist. These are category 1, Drugs or Compounds, which are written in by the indexer as he locates them in the article being scanned, and category 2, the chemical classes to which the drugs or compounds belong. The remaining eight categories make up the Biological Dictionary. All except category 1 are divided into two sections, A and B. Terms in section A are printed on the coding sheet because they are frequently used, and are simply circled by the indexer. Less frequently used terms must be chosen from the dictionary and written in by the indexer in the B section of the coding sheet. Category 1 has no A or B because there are no frequently used Drugs or Compounds. "Drug Actions" are noted in category 3, and general concepts and modifiers which are needed for the clarification of these concepts are included in category 4, which is a catch-all section. Category 4 includes Body Fluids, State and Sex of the living organism being discussed, Pharmacological, Environmental, Nutritional and Toxicity Effects and Processes, Physical, Chemical, and other necessary terms used in indexing papers on pharmaceutical technology. Both methods and apparatus used are of interest to the pharmacists who do research in formulation of drugs. The vocabulary to handle the articles on pharmaceutical technology is very specialized. The remaining categories are Anatomy, Systems and Organs (category 5); Systems, Diseases or Disorders, Symptoms (category 6); Tests and Test Records (category 7); Routes and Types of Administration (category 8); Micro- and Macro-Organisms (category 9); and Fields and Types of Studies (category 10).

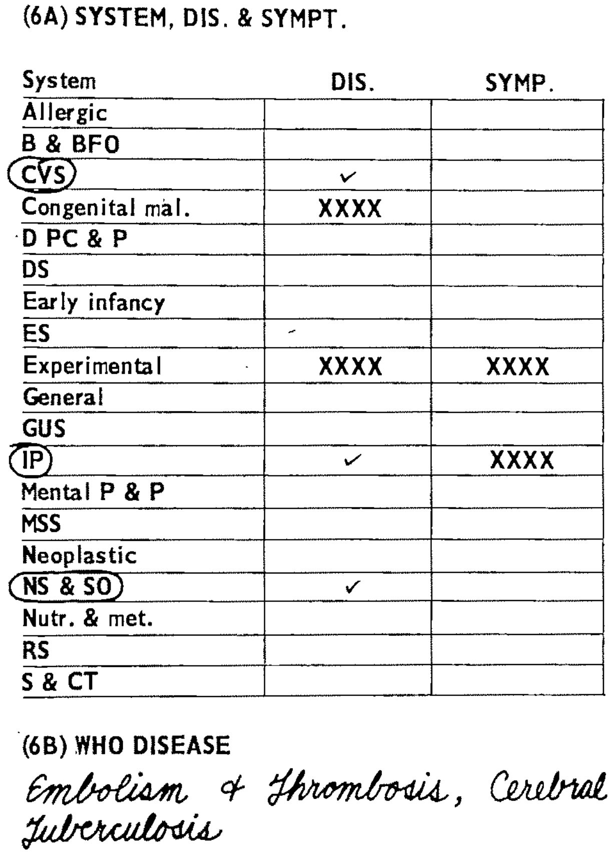
The embryotic dictionary was largely medical in content before the 1960 organization began. The World Health Organization (WHO) Classification of Diseases, 1955 Revision, was used as an authority list with certain modifications in the disease terminology. However, and most important, its classification number was retained. When necessary, disease concepts, recognized after the 1955 Revision, were added to the dictionary within the proper classification group. From the first, the dictionary also included the microorganisms that caused the diseases. *Bergey's Manual of Determinative Bacteriology* was used as an authority list for the bacteria. Other disease-inducing organisms such as viruses, fungi, and protozoa were also included if an infectious disease caused by them was listed. Nonpathogenic microorganisms were added only as they appeared in articles being abstracted.

We decided arbitrarily that the necessary coding terms to index a disease would be the WHO name, the system affected, and the specific organism if known.

Figure 3 presents a tabulation listing System in one column, followed by "Dis." (Disease or Disorder) in another column, and "Symp." (Symptom) in the last column. The disease is written in the blank space below the tabular listing exactly as it appears in the Biological Dictionary. Cerebral Embolism would be correctly indexed as "Embolism and Thrombosis, Cerebral." In the tabular area a circle would be placed around "NS & SO" (Nervous System and Sense Organs) and around "CVS." If the disease were of organic nature, check marks would be placed opposite "NS & SO" and "CVS" in the column headed "Dis." On the other hand, were the disease

1. **Drugs or Compounds**
2. **Chemical Classes**
3. **Drug Actions**
4. **General Concepts and Modifiers**
5. **Anatomy, System and Organs**
6. **System Disorder & Symptom**
7. **Tests & Test Records**
8. **Routes & Types of Administration**
9. **Animals Incl. Microorganisms**
10. **Fields & Types of Study**

FIG. 2. Coding sheet categories



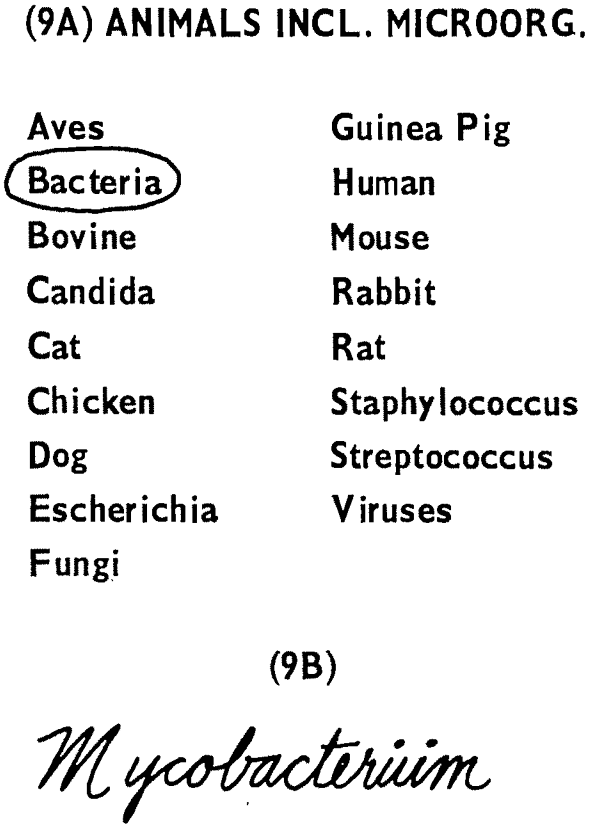
(6A) SYSTEM, DIS. & SYMPT.

System	DIS.	SYMP.
Allergic		
B & BFO		
[✓ CVS]	✓	
Congenital mal.	XXXX	
D PC & P		
DS		
Early infancy		
ES		
Experimental	XXXX	XXXX
General		
GUS		
[✓ IP]	✓	XXXX
Mental P & P		
MSS		
Neoplastic		
[✓ NS & SO]	✓	
Nutr. & met.		
RS		
S & CT		

(6B) WHO DISEASE

Embolism & Thrombosis, Cerebral Tuberculosis

Fig. 3. Disease recording in area 6 of the coding sheet



(9A) ANIMALS INCL. MICROORG.

- Aves
- [✓ Bacteria]
- Bovine
- Candida
- Cat
- Chicken
- Dog
- Escherichia
- Fungi
- Guinea Pig
- Human
- Mouse
- Rabbit
- Rat
- Staphylococcus
- Streptococcus
- Viruses

(9B)

Mycobacteriim

Fig. 4. Area 9 of the coding sheet

attributable to drug administration, check marks would be made in the column headed "Symp." opposite "NS & SO" and "CVS." By making this distinction, it is possible to use the same code for both disease and adverse reactions to drugs.

We carry this distinction throughout the entire Biological Dictionary. All the diseases listed have coding instructions for indicating the cause (normal or drug induced) of illness as discussed in the article being indexed. In infectious disease, for example, "Tuberculosis," the disease name is written in the space below the tabular listing shown in Fig. 3 and IP (Infectious and Parasitic) is circled.

The genus and class (taxonomic) of the causative organism are indexed in area 9 as shown in Fig. 4. "Mycobacterium" is written in area 9B and "Bacteria" is circled in 9A.

In addition to its largely medical content, a few basic modifying terms were included in the original compilation of the dictionary. The dictionary is now in its fourth edition and has been greatly expanded with about 3,500 terms now included. This required a total of about 7,000 IBM cards to prepare and print on an IBM 1403 printer.

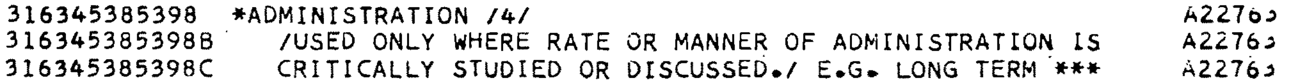
Standardized indexing or coding instructions, to be discussed shortly, are provided for mechanized retrieval of specific biological concepts, including pharmacological activities.

Figure 5 is a partial page from the Biological Dictionary which shows the format we use. From left to right the format shows the random 12-digit number used for mechanized retrieval, the dictionary term itself, a number in slashes, / /, to identify the particular section of the coding sheet to be used, and the sequential number of the term with the proper letter prefix.

Originally, organizing the terms for inclusion in the dictionary required about four to six months of the author's time with the full-time aid of a clerical worker. At the onset, two identical 3 × 5 cards were typed for each term; one set of cards was alphabetized, the remaining set was arranged sequentially by class, according to WHO classification numbers. Pathogenic microorganisms were arranged alphabetically within each family or genus, while general modifying terms were placed in whatever related areas were possible; for example, all toxicity modifiers were grouped together. The other categories were similarly grouped.

The product of the first set of 3 × 5 cards was the alphabetic listing of terms; these were subsequently keypunched and printed to create the dictionary.

The product of the other set of 3 × 5 cards is to be a classified dictionary, now in the process of being key-



```
316345385398    *ADMINISTRATION /4/                                        A22763
316345385398B    /USED ONLY WHERE RATE OR MANNER OF ADMINISTRATION IS     A22763
316345385398C    CRITICALLY STUDIED OR DISCUSSED./ E.G. LONG TERM ***     A22763
```

FIG. 5. Dictionary term, "Administration"

punched. One section, the classified diseases according to WHO, is finished and is already in use.

During the production of the 3 × 5 decks a great deal of time was spent in editing; making cross references to and from synonyms; and, when necessary, appending scope notes to explain or to limit the use of terms.

An example of this, shown in Fig. 5, is the term "Administration" which refers to the administration of a drug. The three asterisks indicate that the modifying terms should be written in parentheses after the term is circled on the coding sheet to bring to the attention of the reader the specific reason why "Administration" was chosen as an indexing term.

Another example is seen in the scope note following the term in Fig. 6: "Chronotropic /affecting the time or rate, especially the rate of contraction; said of nerve fibers that affect the rate of cardiac contraction, the vagus slowing, the sympathetic accelerating/ see Heart & Pulse Rate." In this case the scope note explains the meaning of the term and then refers to the correct term to use in coding the concept under discussion. In many instances, corporate scientists, whose specialty was related to the specific area in which coding terms were being developed, were consulted for clarification (or amplification). For example, when vague or ambiguous terms related to cardiovascular-system effects were encountered, a pharmacologist working on cardiovascular drugs was consulted. The subject specialist decided that the term "Chronotropic," even though used in the literature, was somewhat ambiguous and often loosely used. It was on this basis that "Chronotropic" was explained by definition and included in the general concept of "Heart and Pulse Rate." A scope note is appended to this latter term also.

Throughout the entire early period of organizing the dictionary or thesaurus, a constant effort was made to use the best reference sources available as well as to consult with subject specialists in every instance where a term might be ambiguous in meaning. The choice of the best synonym was also done in this fashion when necessary.

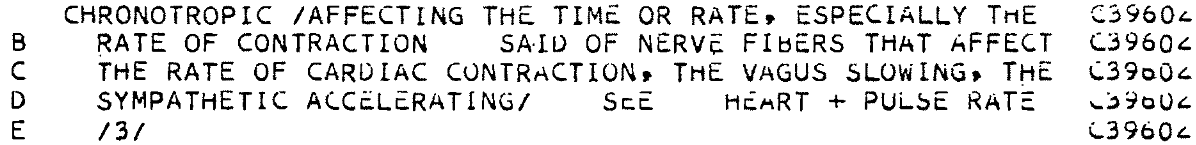
To keep the dictionary manageable in size, many terms were entered generically rather than specifically. For instance, it was decided not to name specific parts of the intestine, such as ileum or duodenum, even if they were specifically named in the article being indexed, but to code only "Intestine /large and small/." "See" references were made from the unused specific terms to the generic term.

This dictionary is open-ended. New concepts can be added and old ones modified at any time. Because it is printed from a deck of key-punched cards, old cards can be pulled and changed or new ones can be key-punched and inserted whenever or wherever necessary.

The 3×5 cards shown in Figs. 7 and 8 depict master dictionary cards. Figure 7 represents a new term, and Fig. 8 a revised term. The new term card has the entry "Depilatory /3/ /%/" typed on it. On the left is the 5-digit accession number (A# 35318) of the abstract where the term first appeared. The number in slashes next to the term indicates that the term is to be entered in area 3 of the coding sheet. The percent sign informs the indexer that this term does not yet have a random number and is to be circled in red when written on the coding sheet.

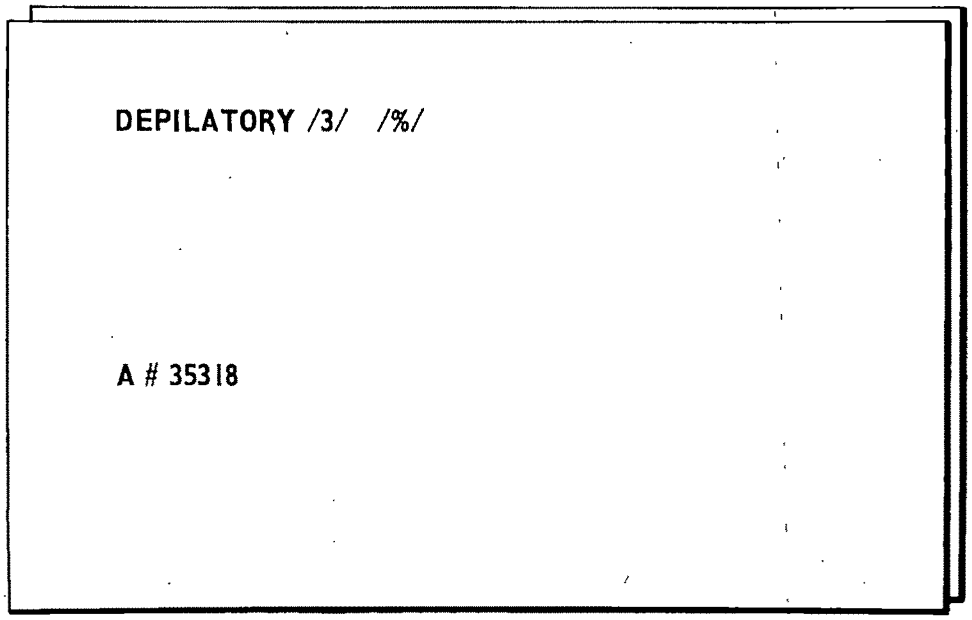
At the end of each week, completed code sheets are scanned by a clerk for red-encircled terms. The number of the abstract is entered in the master 3 × 5 card dictionary. When 10 such numbers are entered, the term is assigned a random number which is the first step in making it machine retrievable. The term is then entered on the 10 master abstract cards in which it had been used and the random number is key-punched in the master search card which represents each abstract. The term is also retyped on a 3 × 5 card with its new random number and an asterisk is placed before the term to indicate that it has a random number and is no longer to be encircled in red when written on the coding sheet.

After the term card has been checked for accuracy a clerk makes a stencil on a Chiang Small Duplicator (17). Each holder of a Biological Dictionary receives a card. If the new term is a disease, a card is also made for each copy of the Classified Diseases Dictionary. The cards for the Biological Dictionary are retained in alphabetical order; those for the Classified Diseases Dictionary are



```
     CHRONOTROPIC /AFFECTING THE TIME OR RATE, ESPECIALLY THE    C39602
B      RATE OF CONTRACTION     SAID OF NERVE FIBERS THAT AFFECT  C39602
C      THE RATE OF CARDIAC CONTRACTION, THE VAGUS SLOWING, THE   C39602
D      SYMPATHETIC ACCELERATING/     SEE    HEART + PULSE RATE   C39602
E      /3/                                                       C39602
```

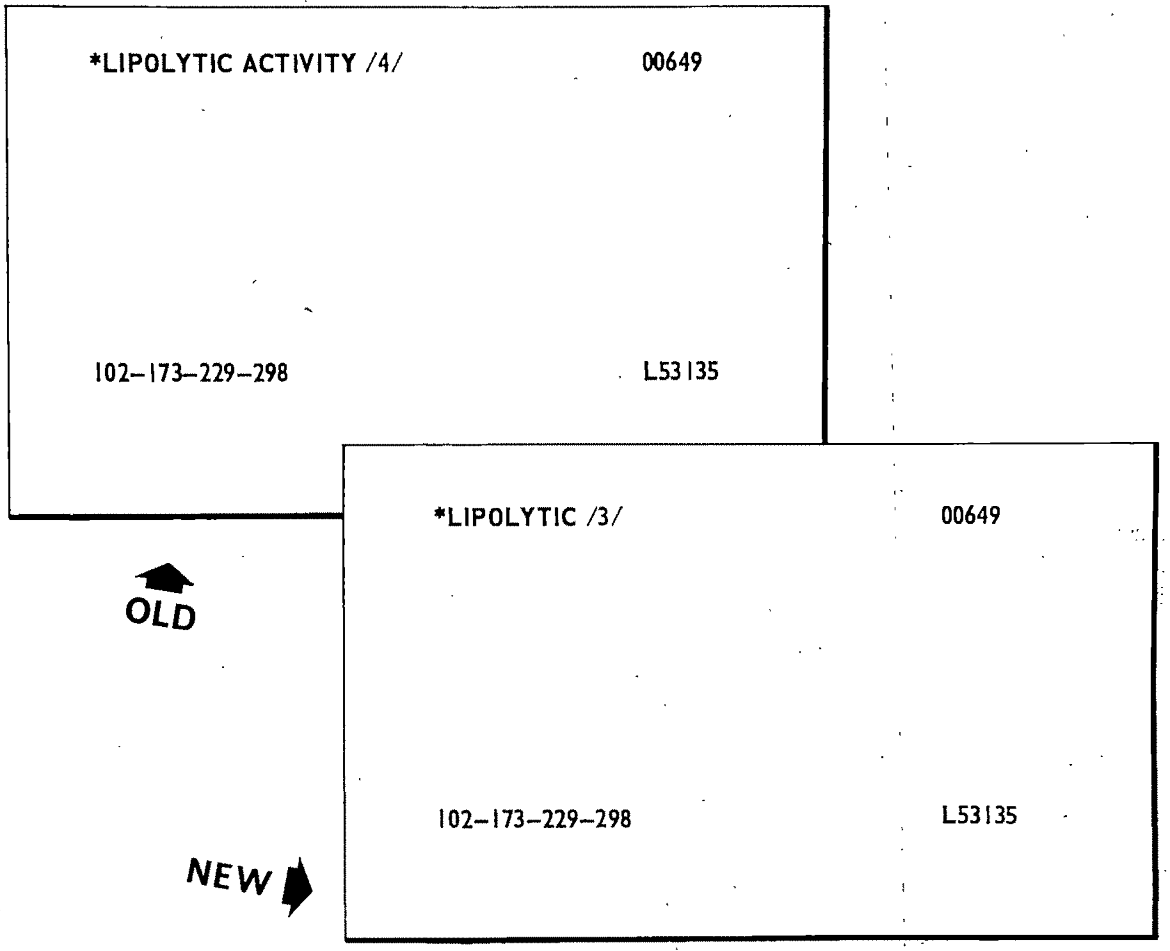
FIG. 6. Dictionary term, "Chronotropic"



DEPILATORY /3/ /%/

A # 35318

Fig. 7. Master 3 x 5 dictionary card, new term, "Depilatory /3/ /%/"



*LIPOLYTIC ACTIVITY /4/ 00649

102–173–229–298 L53135

*LIPOLYTIC /3/ 00649

102–173–229–298 L53135

Fig. 8. Master 3 x 5 dictionary cards, revised term, "Lipolytic /3/"

filed by class number. When indexing, the abstracter-indexer checks both the printed version and the collection of cards representing new terms for the desired concept.

Figures 9 and 10 are flow charts which describe the process of entering a new term (Fig. 9) or making a change in an old term (Fig. 10) for updating the dictionary.

As Fig. 9 depicts, a number of clerical operations as well as various editing steps are required to enter a new term. One step, the assignment of a sequential number, is for the purposes of filing and relocating if cards should be accidentally moved from their normal position. A sequential number is assigned by checking a 5-place log table for the numbers between the two sequential numbers already assigned to the neighbors of the new card. The colored flag instructs the key-punch operator in regard to punching positions.

In the case of a change in terms, for example the assignment of a random number after a term has been used 10 times or a change in meaning, the clerical process is somewhat different. Two 3 × 5 cards are typed with the corrected version of the term. Figure 8 depicts the old and the new master card for the term "Lipolytic." The name was changed from "Lipolytic Activity" to "Lipolytic," and the coding area was moved from area 4 to area 3. A new number appears on this term card in the upper right hand corner. It is the serial number of the random number. This number will shortly be in use since we are in the process of converting our file from cards to magnetic tape. The random number will be replaced by the serial number when the conversion is complete. As shown in Fig. 10, the clerical process is somewhat more complicated and involves more steps for revising terms than for entering a new term. Old punched cards must be clipped to the revised 3 × 5 card for Classified Dictionary key-punching as the instructions for this process are rather complicated. It is easier to give the operator the old cards to use as a pattern than to write new instructions for each change. A change list must also be made and circulated to each holder of the dictionaries. By making the actual correction in his own copy the Information Scientist is alerted to the change in term and will be aware of it in future coding work.

In the Alphabetical Biological Dictionary, terms are in strict alphabetical order even though the term may involve more than one word. Experimental neoplasms

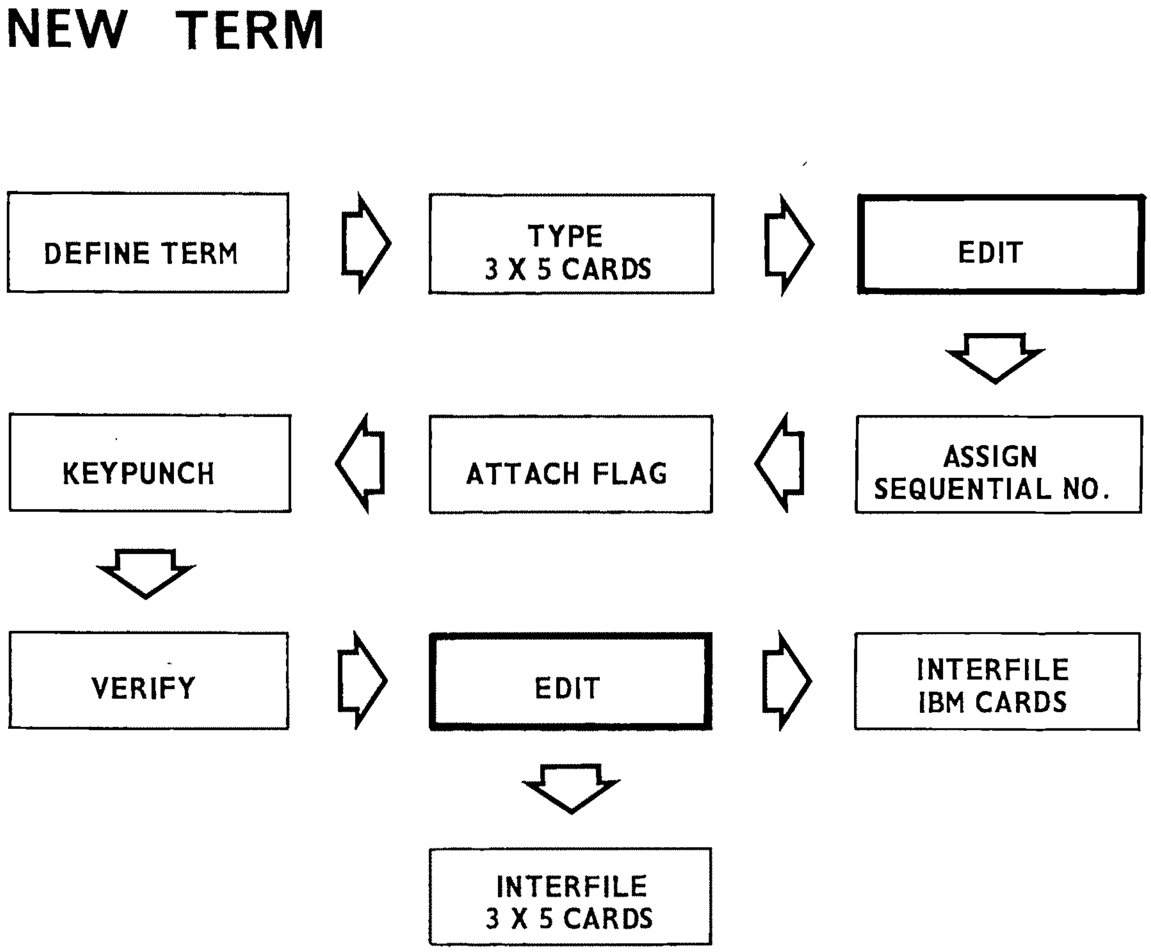


FIG. 9. Procedure to enter new term in dictionary

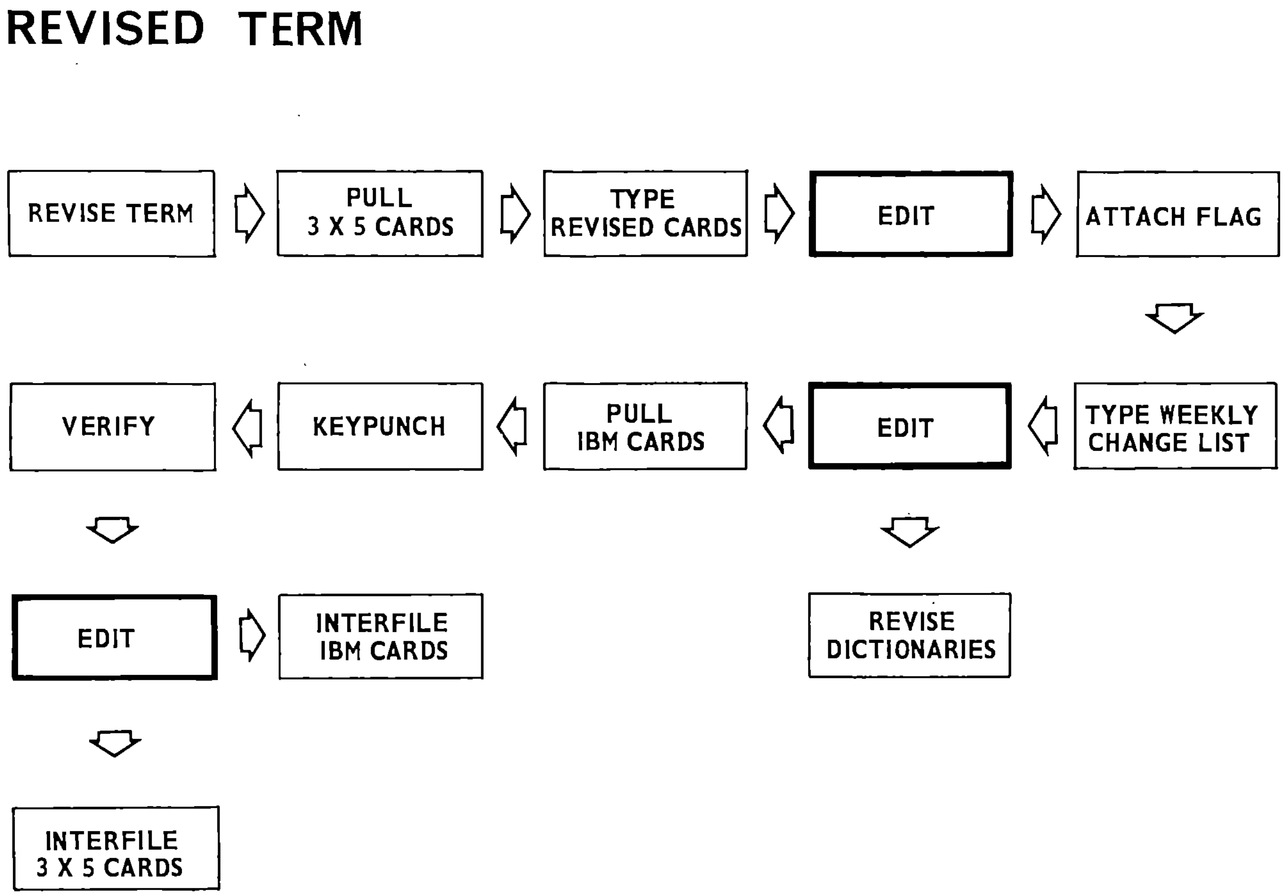


FIG. 10. Procedure to revise term in dictionary

with a number as a whole or part of their name are listed at the beginning of the dictionary. These numbers are arranged in order of increasing numerical sequence irrespective of commas or other punctuation marks. When two or more numerical sequences are identical, letter designations anywhere in the sequence are used as a secondary order of listing.

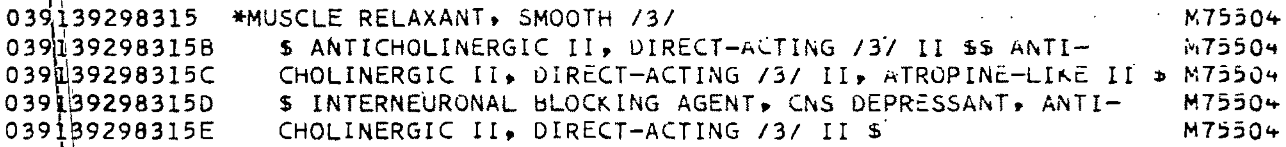
Dictionary entries are followed by obligatory and/or optional terms describing the biological or pharmacological activities reported in the Abbott Abstracts. The descriptive terms for the diseases are followed in every applicable case by the disease classification number assigned to each by WHO. Obligatory coding terms follow the dictionary entries. These terms must be coded when the dictionary term is coded in order to preserve "family" relationships. When the term "Chronic Toxicity" (Fig. 11) is coded the obligatory term is "Toxicity." Optional terms that the indexer may choose are "Pharmacology" or "Clinical Pharmacology." The obligatory use of the term "Toxicity" permits us to search generically for all abstracts in which a toxic effect is discussed. The specific terms so generalized in this case include "Chronic Toxicity," "Acute Toxicity," and "Subacute Toxicity." Every optional term is followed by the symbol "II." If a sequence of optional terms is surrounded by dollar signs and any of the individual terms are used by the abstracter-indexer, all the terms within the dollar signs must be coded.

We update biological and pharmacological activities as more information appears in the literature, and thus we add additional sequences as new concepts are reported. The article being coded determines which sequence is the best to use. The possible sequential sets provided in the dictionary are designed to describe specific and generic concepts.

An example of possible sequences from which the indexer must choose are shown in Fig. 12 under the term

```
036199243348   *CHRONIC TOXICITY /4/                                              C39428
036199243348B     TOXICITY /4/, PHARMACOL. II /10/, CLINICAL PHARMACOL.           C39428
036199243348C     II /10/                                                         C39428
```

FIG. 11. Dictionary term, obligatory entries, "Chronic Toxicity"



```
039139298315   *MUSCLE RELAXANT, SMOOTH /3/                                  M75504
039139298315B    $ ANTICHOLINERGIC II, DIRECT-ACTING /3/ II $$ ANTI-         M75504
039139298315C    CHOLINERGIC II, DIRECT-ACTING /3/ II, ATROPINE-LIKE II $    M75504
039139298315D    $ INTERNEURONAL BLOCKING AGENT, CNS DEPRESSANT, ANTI-       M75504
039139298315E    CHOLINERGIC II, DIRECT-ACTING /3/ II $                      M75504
```

FIG. 12. Dictionary entry, Sequence Choice, "Muscle Relaxant, Smooth"

"Muscle Relaxant, Smooth." Each of the three sequences is enclosed in dollar signs.

To summarize this brief description of the preparation, control, and maintenance of Abbott's Biological Dictionary, it can be stated that our experience has shown that firm control of any dictionary or thesaurus must be maintained. When reliance is placed solely on indexing terms chosen from the article being indexed, it is inevitable that a loss of pertinent material will occur when a search is being made. If no thought is given to family or generic entries each time one of the family is being indexed, the loss will be even greater. Non-thesaurus, uncontrolled indexing is very wasteful and unreliable; therefore, one should give careful consideration to the format of a thesaurus or dictionary before adapting it for use.

The appendix shows the special marks used to facilitate instructions to indexers and to accommodate IBM printouts as well as acceptable abbreviations.

Appendix

Special marks are used to facilitate instructions and to accommodate IBM printout.

Part 1—Special Punctuation

SYMBOL	DESCRIPTION
/%/	Term does not have a random number; encircle in red on the coding sheet.
*, **	Term has a random number. (A single asterisk indicates that a term has a random number. Two asterisks means that the term has a random number and is also generic, e.g., *Mycobacterium, **Bacteria.)
***	Three asterisks following a term mean that explanatory notes are to be inserted in parentheses after the term.
/1/, /2/, /3/, /4/, /5/, /6/, /7/, /8/, /9/, /10/	Designation of appropriate areas on coding sheet.
/ /	Parentheses or brackets.
$ $	Enclosing coding sequences and unique instructions.

Part 2—Abbreviations. The following accepted abbreviations are used in the dictionary.

ABBREVIATION	DESCRIPTION
A#	Abbott Abstract Number
ANS	Autonomic Nervous System
B & BFO	Blood & Blood Forming Organs
CNS	Central Nervous System
CVS	Cardiovascular System
CONGENITAL MAL.	Congenital Malformations
D PC & P	Deliveries, Pregnancy and Childbirth, and Puerperium
DS	Digestive System
DIS	Disease or Disorder
ECG	Electrocardiogram
EEG	Electroencephalogram
EN	Endogenous
ES	Endocrine System
EX	Exogenous
GS	Genital System
GUS	Genito-Urinary System
II	If Indicated
I.A.	Intra-Arterial
I.M.	Intramuscular
INHIB	Inhibitor
I.P.	Intraperitoneal
IP	Infective & Parasitic
I.V.	Intravenous
LIMIT	Limited to
MAO	Monoamine oxidase
MENTAL P & P	Mental, Psychoneurotic & Personality
MSS	Musculo-Skeletal System
MUSCLE, TEND. & FAS.	Muscle, Tendon & Fascia
NS & SO	Nervous System & Sense Organs
NERVES & PERIPHER. GANG.	Nerves & Peripheral Ganglia
NOS	Not Otherwise Specified
NUTR. & MET.	Nutritional & Metabolic
RS	Respiratory System
S & CT	Skin & Cellular Tissue
SC	Subcutaneous
SPECIFIC	Specific Term
SYMP.	Symptom
US	Urinary System

References

1. HOLM, B. E., and L. E. RASMUSSEN, Development of a Technical Thesaurus, *American Documentation,* 12 (No. 3):184–190 (1961).

2. BERNIER, C. L., and K. F. HEUMANN, Correlative Index III, Semantic Relations Among Semantemes—The Technical Thesaurus, *American Documentation*, 8 (No. 3):211–220 (1957).

3. TAUBE, M., The Preparation of Manual Dictionaries of Association. In *Coordinate Indexing*, Vol. II, Documentation Inc. (1954).

4. TASMAN, P., Literary Data Processing, *IBM Journal Research & Development* (1957).

5. JOYCE, T., and R. M. NEEDHAN, The Thesaurus Approach to Information Retrieval, *American Documentation*, 9 (No. 3):192–197 (1958).

6. BERNIER, C. L., Correlative Indexes II: Correlative Trope Indexes, *American Documentation*, 8 (No. 1): 47–50 (1957).

7. LUHN, H. P., A Statistical Approach to Mechanized Encoding and Searching of Literary Information, *IBM Journal Research & Development* (1957).

8. LUHN, H. P., Potentialities of Auto-encoding of Scientific Literature, IBM Research Report RC-101, May 15, 1959.

9. Armed Services Technical Information Agency, *Thesaurus of ASTIA Descriptors*, First Edition, May 1960; *ibid*, Second Edition, April 1964.

10. *Chemical Engineering Thesaurus*, American Institute of Chemical Engineers, New York, 256 pp., 1961.

11. *Thesaurus of Engineering Terms* (First Edition), Engineers Joint Council, New York, May 1964.

12. WALL, E., Final Report—Final Revision of Thesaurus of ASTIA Descriptors, Armed Services Technical Information Agency, Engineers Joint Council, AD-278168. 19 pp., August 6, 1962.

13. U.S. Bureau of Reclamation, *Thesaurus of Descriptors; A List of Key Words and Cross References for Indexing and Retrieving the Literature of Water Resources Development* (Tentative Ed.), Denver, 1963.

14. *U.S. Navy Bureau of Ships Thesaurus of Descriptive Terms and Code Book* (Second Edition), Washington, March 1965.

15. *American Petroleum Institute Subject Authority List*, New York, 1964.

16. *COSATI Subject Category List*. Federal Council for Science and Technology, Executive Office of the President, Washington, D. C. 55 pp., December 1964.

17. Chiang Small Duplicator, 531000 Juniper Road, South Bend, Indiana 46637.

18. SOUTHERN, W. A., Mechanized Processing and Retrieval of Bio-medical Information, *Methods of Information in Medicine*, 1:16–22 (1962).

Classification Systems and Their Subjects[1]

A General Analysis of Different Kinds of Classification Systems Characterized by Different Types of Subject

The purpose of this paper is to analyze different types of classification systems characterized by different types of subjects. The most important types of subjects in documentation are *documents* and *terms,* but other systems for other subjects are discussed. Systems of science are dealt with as an introduction to document systems. Every system ought to have a structure that will fit the subjects. The hierarchies for documents and terms are not the same. If we try to classify a category of subjects in a system made for another category of subjects (for example, classifying terms or technical products according to a document system) we will always meet with difficulties.

We should not be bound for the future to now existing systems, but for each type of subject we have to create the best possible system or variant. Our aim should be to get a system of systems, where every individual system as far as possible—without distortion of its primary function—has common features with the other systems. Principles for designing a modern universal document system based on concepts and adapted to technology and a universal term system are advanced.

EJNAR WÅHLIN

Stockholm

By way of introduction I may quote a statement from the FID/CR Conference at Elsinore in the autumn of 1964 (1):

> The Conference listed as its aims:
>
> the improvement of existing classifications, including work on methods for construction of thesauri and related tools;
>
> the achievement of better design in new classifications;
>
> the exploration and implementation of compatibility among classification systems and thesauri, including standardized vocabularies;
>
> the convertibility of the records of material indexed in one system into another;
>
> and the study of the interaction between classification systems and computer technology in the process of system analysis and programming.

These recommendations have the great merit that they go deep into the core of the problem and may therefore be considered fruitful. Most of these points are discussed in this report.

The words "classificaton," "classification systems," and "systems" occur in all five paragraphs above. It is important to know exactly what kind of systems are referred to here.[2]

With regard to the scope of the system we talk of *universal systems,* comprising all fields of knowledge, and *specialized systems,* covering only a limited field. The general and traditional structure of classification systems is the *hierarchical* (tree-structure), but among specialized systems *faceted* systems are a particular type. These differ from the traditional *discipline systems* not only in structure but also in being entirely based on concepts.

Examples of adjectives used to denote the meaning of the word classification in special contexts are mentioned by R. Mölgaard'H (in the proceedings of the Elsinore Conference 1964): "informative, topological, dynamic, synthetic, faceted, hierarchical, factor-analytical, natural, arbitrary, general, special, and topographical."

The most important and fruitful approach seems to me, however, to characterize the systems according to *the subject of the system,* i.e., what the designer of the system wanted to arrange and classify, but very little attention seems to have been paid to this problem. We find in the literature that some experts, in writing about classification systems, have classification of *knowledge concepts,* or *ideas* in mind, others classification of *documents,* while sometimes the subject is *terms, things,* or other phenomena. I shall not maintain that some of them are right, others wrong. In fact, different kinds of classification systems with different types of subjects may be necessary in documentation.

The types of systems I wish to deal with are all "gen-

[1] Financial support by The Swedish Council for Building Research and The Swedish Council for Technical Research.

[2] Concerning the nomenclature in the sequel, I shall consider "classifications" and "classification systems" as synonymous. "System" is sometimes used in a wider sense to express contexts which cannot be presented as written schemes, but the word is not used in the sense of "information system."

eral," in the sense that they are not restricted to a special subject field, and they are all of importance for documentation. They are all types of *systems in existence,* having been devised by different professional groups for different practical requirements.

I shall first discuss *systems of science:* several systems of this kind have been devised and here we have a background to the document systems. I shall not discuss systems of knowledge as I do not think that there exists any reliable system comprising all knowledge for the specific purpose of classifying knowledge, but I shall deal with *systems for documents* based on their knowledge content. I will interpret the heading Mathematics in a document system as "Documents on Mathematics" and the subheadings as "Documents on Algebra," etc. There are other systems (even if not called classification systems) where the heading Mathematics has another significance, e.g., "Teaching of Mathematics" (in a timetable) or "text on mathematics" (in a handbook).

Even if a document system is *based on concepts,* I will avoid presenting it as a system for concepts, because the purpose is not to classify concepts, but to classify documents.

A universal system for concepts (made with the purpose to classify concepts) does not exist and is not easily made because there doesn't exist any definite collection of concepts from which we can pick up concepts and arrange them.

On the other hand concepts are usually expressed by terms and we can talk of *systems of terms* as we really have a definite stock of terms. The terms here are the subjects and should determine the structure of the system.

Also "practical systems" used in industry and commerce for registration purposes will be discussed. A *system for products* should be designed according to the structure of the collection of products uninfluenced by documentation aspects.

The essential point in this presentation is to show how and why these systems have come about, how they have developed, what role they play or should play in documentation, and to indicate principles on which we can create for every type of subject a new system designed to fulfill its function in the best possible way, all systems forming a *system of systems* in which every individual system—as far as possible without disturbing its primary function—has common features with the other systems.

• 1. The System of the Sciences and the Systems within the Sciences

A. SCIENCE AND SYSTEM

According to Albert Einstein, science is the striving, with the aid of thought, to arrange the observable phenomena of our world into, as far as possible, a coherent *system,* an attempt to reconstruct our existence with the help of the formation of *concepts.* The main purpose of documentation is to classify scientific literature and make it accessible. The structure (or system) of science can then not be without significance.

The order between the individual sciences should not be a question of prestige. Only by arranging a science in a certain series is it possible to know what is *before,* what is *within,* and what is *behind,* and thus to get a definition of each science.

The history of the system of the sciences is described by, among others, Bliss (2) and, in Swedish, by Allen Vannerus (3). The latter has made very thorough studies and presents a system that is still of interest. He defines the "systematics of the sciences" as follows:

> What must be done is to bring the sciences, this characteristic and highly important element in our spiritual culture, under the form of a *system* in order to obtain an insight not only into their multiplicity but also into their affinities.

In older times the natural starting point for philosophers was the human mind rather than nature. To quote from Francis Bacon (4):

> The parts of human learning have reference to the parts of man's understanding, which is the seat of learning: History to his *Memory,* Poetry to his *Imagination,* and Philosophy to his *Reason.* . . . Thus I have made, as it were, a small globe of the intellectual world.

During the 18th century most natural sciences developed rapidly, but the continuity was still obscure. Fossils of plants and animals in rock were sometimes interpreted by scientists as products of "Vis plastica," i.e., a supposed power of nature to reproduce living beings in stone. The history of creation and the chronology of the Bible were for most scientists a reality. Obviously the time was not ripe for a coherent system of science.

In the 19th century—in conjunction with the development of the theory of evolution—a picture became clear of the relations between the sciences which in broad outline still holds today and which implied a radical reversal of the earlier order of sequence between the sciences. This very important reversal seems not to have been observed in books on the history of the system of science. It is moreover paradoxical that during that century, when it first became possible to speak of a general picture of the sciences which could be accepted in the whole western world, the interest in the system of the sciences began to die out both among philosophers and librarians. H. E. Bliss is here an exception, as de Grolier (5) points out:

> His principal effort was directed to one point; to draw bibliographic classification nearer to what he termed "the scientific and education consensus"—the scientific and pedagogic order of subjects of study. His work from this standpoint has historical importance.

A documentalist now engaged on the relation between the sciences is the Pole, Z. Dobrowolsky (6), who wants to create a new "encyclopaedic classification."

This would be important, not only for the development of documentation but also for the furtherance of scientific research by furnishing a basis for rational organization of research within all fields of knowledge and on an international level.

It is true that documents nowadays cannot as easily as before be fitted into traditional folds, since scientific progress has meant that different sciences are to an ever increasing extent woven together and make use of one another. But this does not signify that a system is lacking.

B. Systems within the sciences

Apart from the structure of the whole field of science (the system of science), most sciences have developed rather detailed and fixed systems or structures of their own. Linne's "Systema Naturae" is an example, perhaps not quite up to date, but in his time bringing order out of chaos. The systems of the chemical elements, of atoms and electrons, of the stars and the galaxies, of the hereditary factors, etc., are other well known systems, forming a basic layer in the individual sciences and in the whole field of science as well.

C. Main sequence in science

A division of the entire field of science into only three main groups—which can scarcely be questioned—becomes clear from the following:

1. In the beginning there existed only lifeless matter and energy. The processes which went on are described in the sciences of mechanics, physics, chemistry, astronomy, geophysics, etc., with mathematics as basic science. The heavenly bodies and our earth appeared.
2. The origin of life signifies the start of a new epoch in natural history with a new series of phenomena and concepts. Living matter, plants, and animals evolved.
3. When some higher mammals developed into man, still another phase starts. Language arises, nature is exploited, societies and cultural activities come into being, and the stock of concepts was rapidly increased.

We have here a simple tripartite principle:

Matter (i.e., inorganic matter and energy)
Life (i.e., living matter and physical life)
Culture and society (man as an individual and in society)

This principle corresponds with that advanced by the British librarian, James Duff Brown (7), who at the beginning of this century suggested the simple series:

Matter—Life—Mind—Record

Figure 1 shows two philosophical systems from the first half of the 19th century and the library systems of Bliss and Brown constructed a century later which all—by and large—exhibit the sequence mentioned above. In contrast thereto, UDC (and Dewey) exhibit quite another structure representing a picture antiquated already in Dewey's time, and this also applies to many library systems used today.

Nowadays—with our strong trend toward specialization—these questions are considered a little old-fashioned. The field of knowledge is regarded by many documentalists as a coherent mass without distinguishable dividing lines; or the only way to divide it is according to some of the well-known traditional library systems. Only very occasionally is a new voice heard. It was therefore a strange coincidence that, at the time of writing this paper, the author happened to read an article in a Swedish newspaper entitled "Science with or without method." Dr. Boris Tullander (8), Lecturer in Economic Methodology at Uppsala University, writes, though not with a thought for documentation:

> There are at least three areas of our existence which are sharply distinguished. One may call them (1) the sphere of the material—mechanical relations, (2) the sphere of the organic—biological relations, (3) the sphere of the psychological—sociological relations. These relations, or "causa nexi," are geared to one another but they also represent a distinct advance and one immediately recognizes the differences.

Coming down to a lower level the principles for further division are not so self-evident.

A division of the sciences not consistent with that above is advocated by Maurice Korach (Hungary) in an interesting analysis (9) of the nature of the sciences:

Pure sciences	Natural sciences	Technical sciences, technologies
Mathematics	Zoology	Agricultural technologies
Physics	Botany	Industrial technologies
Chemistry	Mineralogy and petrography	———
etc.	Astronomy	
	etc.	

This division seems to aim at the general character of the sciences and not at their knowledge content (e.g., Zoology and Botany between Chemistry and Petrography). Perhaps we have to admit that the system of science can be regarded from different aspects. The documentation aspect, however, seems better justified by the scheme advocated above.

D. Disciplines or concepts

To set up a fixed and detailed scheme, with the traditional disciplines as building blocks, has its difficulties even if we are agreed on the principle of their sequence. There have been successive changes in the composition of the sciences. Philosophy originally comprised nearly all science, but later had to relinquish bit by bit. The limits between mechanics, physics, and chemistry are

	COMTE	SPENCER	BLISS		BROWN	UDC	
		Logics	A.	Philosophy			Generalities
	Mathematics	Mathematics		General Science		0	Bibliography
I	Astronomy	Mechanics		Logic, Mathematics, etc.			Libraries, etc.
	Physics	Physics	B.	Physics	Matter		Philosophy
	Chemistry	Chemistry	C.	Chemistry		1	Ethics
		Astronomy	D.	Astronomy			Psychology
				including Geology,			
		Geology		Geography		2	Religion
							Theology
	Physiology	Biology	E.	Biology		3	Social sciences
II			F.	Botany	Life		Law
			G.	Zoology		4	Philology
			H.	Man			Linguistics
				Physical		5	Pure sciences
				Medicine			
							Applied sciences
		Psychology				6	(Medicine, Technology)
	Sociology	Sociology	I.	Psychology	Mind		
				Social			Arts
III				Education		7	Entertainment
			K.	Sociology			Sport
				Ethnology			Literature
				Human Geography		8	Belles lettres
				Travel and Description			
			L/O.	Social-Political			Geography
				History		9	History
			P.	Religion and Ethics			Biography
			Q.	Social Welfare,			
				Applied Ethics			
			R.	Political Science			
			S.	Law			
			T.	Economics			
			U.	Useful Arts, Industries			
				Trades			
			V.	Fine Arts, Philosophy			
			W/Y.	Literature and Language	Record[1]		
			Z.	Bibliography			

[1] Record = Literary forms, History, Geography and Biography

FIG. 1. Two philosophical systems from the first half of the nineteenth century and two library systems constructed a century later compared with another and with UDC

debatable. Atomic and nuclear physics have broken away from classical physics to form a new fundamental physics related to optics and the theory of electricity. Space science is another example and similar conditions prevail over the entire science register. When we come to the applied sciences, it is difficult to draw a frontier with pure science and no generally accepted subdivision of technology exists. Within sociology the difficulty is still greater, not to speak of psychology.

In the introduction we distinguished between traditional discipline systems and concept systems. By subdividing the total field of science *according to concepts* one obtains a basis for a system free of the traditional boundary lines.

This will not imply a revolution in the sequence, as one can follow in broad outline the sequence presented in Table 1 and thus make use of the already disentangled relations. But now we are more free to follow the principle of introducing new basic concepts in such a way that each new group can make use of earlier concepts and no group need make use of concepts from subsequent groups. One can also, within the pure basic sciences, obtain a close agreement with the system for units of measure, magnitudes, and physical "dimensions" as shown in Table 1.

A more extensive sketch of the sciences and their objects, embracing the introductory groups, is shown in Table 2.

TABLE 1. The Pure Sciences, Their Magnitudes and Basic Units of Measure

Science	Magnitudes[1]	Dimension	Units of measure[2]
Arithmetic	Number	0	—
Algebra and analysis	Mathematical quantities	0	—
Geometry	*Length*	1	m
	Area, etc.	1^2	m^2
Chronology	*Time*	t	s
Kinematics	Velocity	$v = 1/t$	m/s
	Acceleration, etc.	$a = 1/t^2$	m/s^2
Statics	*Force*[3]	F	kp or Newton[4]
	Moment of a force	$F \times 1$	kpm
	Pressure, etc.	$F/1^2$	kp/m^2
Dynamics	*Mass*	m	kg
	Kinetic energy, etc.	$m \times v^2$	$kg\ 1^2/t^2$

[1] Basic magnitudes in italics.

[2] m = meter, s = second, kg = kilogram, kp = kilopond (1 kilopond is the gravity of 1 kg under certain conditions).

[3] In statics force can be considered as a fundamental magnitude; but if, as in the mks system, mass is a fundamental magnitude, force will be derived by Newton's law: force = mass × acceleration ($F = m \times a$).

[4] 1 Newton is that force that will give the mass of 1 kg an acceleration of 1 m/s^2.

TABLE 2. The Pure (Exact) and the Natural Sciences and Their Respective Concepts

I. The pure and inorganic material sciences	
A. Pure nonmaterial sciences (= exact sciences)	*Concepts*
Mathematics	Number and mathematical quantities
Geometry	Space[1]
Chronology	Time[1]
Kinematics	Motion
Statics } Mechanics	Force
Dynamics } Mechanics	Mass[2]
B. Pure material sciences	
Physics	Matter
Chemistry	and Energy
C. Sciences of the universe and the earth (= inorganic natural sciences)	
Astronomy	Universe
Geophysics	The earth
Geology	The solid surface of the earth
Hydrology. Oceanography	The water on the face of the earth
Meteorology	The atmosphere
Space science	The space around us
D. Technology	
II. The biological sciences (= organic natural sciences)	
A. General biology	Life
B. Botany	Plants
C. Zoology	Animals
D. Anthropology. Medicine	Man
III. ———	

[1] The concepts space and time are not the same as in Ranganathan's space and time facets; they do not embrace geography and history.

[2] Mass is not to be mixed up with matter or material. Mass is ideal matter, characterised only by weight and inertia.

The more dynamic the development in our stock of knowledge, the more important it is, in the construction of a system of sciences, to build on simple, permanent, fundamental concepts. Concepts such as number, force, mass, energy, heat, atom, molecule, metals, wood, electricity, cell, heredity, sex, organs of sense, etc., stand through all times whatever the developments in society and technology. Changes in the sequence of the fundamental concepts are—after Newton and Darwin—rare. Einstein's theories may have led to a new conception of the most elementary basic concepts, but this is hardly of practical significance for documentation.

For dividing the field of technology this principle is of importance by creating a firmer ground for classifying those parts of technology which constitute applied science. We shall come back to this question in Section 2.

To extend the system of sciences to all areas influenced by the activity of man is exactly the purpose and the aim of science, because science is the knowledge of systematical relations between concepts, even if we cannot construct a hierarchy comprising all details of science.

• 2. Document Systems

By document systems is meant systems designed for classification of documents, whether for arrangement on the shelf or for organization of card indexes or printed indexes, each document being classified according to its knowledge content. As earlier intimated we can, when trying to classify a certain piece of knowledge, use different principles for different purposes. Therefore we have a more pragmatic approach if we talk of classifying documents according to their knowledge content than if we talk of classifying knowledge.

2.1. HIERARCHICAL DOCUMENT SYSTEMS

In this category we can distinguish between universal and specialized systems. The field of interest is here divided from the above in classes, subclasses, etc., down to a certain level. For each document we try to find one or more places where the document as a whole has its domicile.

2.1.1. UNIVERSAL DOCUMENT SYSTEMS

This is the category of the well-known library systems as Dewey, UDC, etc. The systems used in the book trade also belong here.

A. *Main Structure*

The main structure of these systems is generally influenced by the division into disciplines as adopted in the universities. The production of scientific literature followed in old days this division in broad outline and could therefore be easily fitted in the pattern.

The similar organization of studies at the occidental universities meant that the building blocks in the schemes of different countries were roughly the same, which facilitated an international diffusion of certain systems.

The traditional sequence, starting with philosophy, religion, and sociology, was accepted by Dewey and was characteristic also of the UDC based on Dewey's scheme. For the applied sciences one special main group was introduced.

As time went on, the document systems acquired an increasingly detailed structure, especially within the practical fields, and the connection with the traditional disciplines was partly dissolved.

The subdivision of the applied sciences and other practical fields was generally based on the differentiation into industries and professions and did not stand in any carefully thought-out correspondence with the system of sciences, but partially duplicated the latter. This was a natural arrangement, for within other fields—mathematics, biology, law, art, etc.—it was to a large extent particular professional categories that corresponded to the document groups, both as authors and readers.

B. *Principles of Subdivision*

We will first investigate the principal structure of UDC—as a representative for this category of systems. The main classes have already been dealt with. On lower levels UDC displays in some parts real generic divisions, e.g., in language, botany, zoology. In other parts we have non-generic but strictly hierarchical divisions, implying that the subclasses are equal and parallel. This is valid, e.g., for anatomy, for physics (the traditional subdisciplines), for many parts of technology, etc.

To a great extent, however, the subclasses correspond *to different aspects* of the class, the aspects having their domicile in other parts of the system. An example:

631.	Agriculture, farming in general. Agronomy
.1	Farm management
.2	Farms and farmyards
.3	Agricultural implements, tools and machinery (e.g., ploughs)
.4	Soil science (e.g., physical properties)
.5	Growing, cultivation methods (e.g., ploughing)
.6	Rural engineering (e.g., drainage)
.8	Fertilizers. Manuring (e.g., nitrogen fertilizers)
.9	Other topics

Most subclasses are applications of one special science or technique (economy, building, machinery, geology, etc.). We have now definitely left the "tree of knowledge" and the connection with the systematic of sciences. The system is a tool for collecting—in a collection comprising more than agronomy—all documents written for agronomists. The codes could also have been derived by combining 631 with other UDC-numbers.

C. *Precoordination of Concepts*

It may be reasonable to ask how in a document system we can introduce a limited number of concept combinations from the very great number of possible combinations, and if it is possible to design rules regulating which concept of two should have preferences. Is a universal document system perhaps on the whole an absurdity?

First, we have to consider that a document can often (depending on the size and the character of the collection) be classified only with the help of a single concept.

Second, to a certain extent, precoordinated concept combinations can be introduced and be a valuable feature.

Not all combinations of concepts have any real importance; the number studied and written about in the literature is in fact limited in relation to the number of possible combinations, even if it is large. A specialized documentalist will recognize certain types of combinations which often recur.

If a document deals with the concepts *a* and *b* in relation to one another, and both of them are represented in the system, the document can naturally be classified both under *a* and *b*. But if *ab* is an important combination, which has caused a great production of literature, there may be reason to introduce this combination in the scheme.

Let us take an example that is applicable both for a universal collection and for a special collection (e.g., a building trade library), but perhaps *not* for a specialized research institute.

On a purely generic basis the literature on concrete would be classified according to type of concrete. But as the literature is dominated by normal concrete made of Portland cement and ordinary stone material, it is perhaps more appropriate to form a system such as the following:

Concrete
  Manufacture of concrete
  ———
  Properties of concrete
    General
    Mechanical properties
      Strength
      Hardness
    ———
    Physical properties
    ———
  Special types of concrete

Combinations of concepts such as *concrete : strength* thus have their given place in this arrangement. The combination *strength : concrete* should in such case not be used.

In many cases the priority of the concepts *a* and *b* can be questioned but in other cases one sequence is definitely to be preferred. The building library will surely prefer to bring together all documents on concrete instead of bringing together all documents on strength. "Stronger" concepts like special materials, things, etc., will have preference over "weaker" concepts, e.g., properties.

We thus conclude that *concept combinations, if carefully selected, can be valuable features* in a document system, but it seems very hard to make such a selection if the structure of the system is not logically built up.

D. *Postcoordination of Concepts*

Even if the system includes the most important combinations today, new important combinations will come forward and we must have a method for coding them. How does UDC function in the case of unforeseen combinations? Let us first see how UDC is presented by FID. We quote from the General Introduction to the trilingual edition (my italics) (10):

> The Universal Decimal Classification (UDC) is a scheme for classifying the whole *field of knowledge*....
>
> a classification in the strictest sense, depending on the analysis of idea content, so that related *concepts* and groups of concepts are brought together.
>
> . . . an integrated pattern of correlated *subjects*. . . .
>
> . . . constructed on the principle of proceeding from the general to the more particular by the (arbitrary) division of the whole of human *knowledge*. . . .

We also read that:

> it should not be regarded as a philosophical classification of *knowledge*.
>
> a practical system for numerically coding *information*, so designed that any *item*, once coded and filed correctly, can be readily found from whatever angle it is sought.
>
> . . . the introduction of an auxiliary apparatus of connection and relation signs, lacking in the original Dewey system, has made the UDC really universal, in the sense that it permits almost any desired combination and modification of basic numbers to denote the most complex *subjects*.

It may be true that the system permits coding of practically any combination of concepts by linking two concepts with the colon sign or with the use of the auxiliary tables. But is this coding of such a nature as to provide sufficient guidance for searching? The concepts to be combined for forming complex subjects are to a large extent already combinations of concepts (type *ab* or *abc*). The same concept is for that matter often located in more than one place in the system.

Can one not often express a certain combination of concepts in different notational ways. Perhaps *ab* : *lm* has the same signification as *al* : *bm*, etc., but how is one to choose the right entrance, especially if a concept is to be found in more than one place in the system?

This combining of concepts, which cannot in themselves be regarded as simple, fundamental concepts selected on a strictly logical principle, leads to a fairly complicated procedure. Is this not a compromise between two logical principles? The principles are as follows:

1. Classification with the aid of a document system based on both generic and other strictly hierarchical structures and moreover on certain precoordinated structures (*ab* and *lm*) and

2. Indexing with the aid of a number of equivalent terms (*a*, *b*, *l*, *m*).

E. *How to Get a Modern Universal Document System?*

The first condition for a new universal system is that it should be linked up with the scientific pattern which has been generally accepted for more than 100 years (cf. Bliss) and the second, that it should be designed on the basis of fundamental concepts instead of traditional headings.

These questions have been dealt with above with respect to the system of science. A proposal for the main structure of a document system based on these principles has been made by the author (11). Here the possibility is also advanced of different universal document systems for different main fields of activity, each system being universal but designed from a special aspect. (See Section G.) The proposal mentioned is especially intended for technology and called *TUS*, i.e., *Technological Universal* (*Document*) *System*. The main groups are (cf. Table 2):

1.	Abstract concepts	I A
2.	Force. Energy	I B
3.	Matter. Materials	I B
4.	World	I C
5.	Life. Man	II
6.	The individual. Humanistic fields	III
7.	Society	III
8.	Material culture	III
9.	History. Geography. Biography	III

The principle behind the main groups is explained in Section 1, p. 203. One important factor for a system intended for technology is naturally the principle for the division of technology. This question has here been solved in such a way that all technical subjects, which can be related to fundamental concepts belonging to those parts of the system which are dependent on the system of science, are located in connection with these fundamental concepts; while those parts of technology which are best characterized by their purpose, e.g., Transport, Building, etc., are collected in one main group, called "Material Culture" or "Functional Material Culture." This group will contain subjects of primary interest to man and society (automobiles, buildings, etc.), while that which lies behind (the engine, installations, theory) is considered as secondary and appears in connection with the fundamental concepts. Strictly speaking, two separate bases for classification have been used:

	A	B
(1)	Applied sciences	Nonapplied sciences
(2)	Not of primary interest to man and society	Of primary interest to man and society

These two principles of classification, however, lead to a great extent to the same result. We find the same tendency to division of technology into two parts if we study the titles of the main branches of technological teaching:

A. By science: mechanical, chemical, electrical (etc.) engineering

B. By purpose: housebuilding, shipbuilding, aviation, mining, etc.

TUS has in the first place to be judged from the viewpoints of special fields of activity, which only collect documents for their own needs. Every field has its own main number in the system, under which the specific documents will be collected, but has to spread out other documents over the whole system as far as possible. Different applications of the system for different types of document collections are discussed in the sequel.

These principles for designing a universal system were advanced as early as 1949 at a conference on Building Documentation in Geneva (12). In the last years it has been possible to go a little further on this road and TUS has now been constructed in detail as regards those parts which are of interest to the housebuilding field. It has been adopted for a bibliography comprising 25 years of the Swedish journals *Arkitektur* and *Byggmastaren* (13). By way of experiment the whole of UDC in its trilingual edition has also been rearranged under TUS headings, which permits a considerable concentration of related fields of knowledge which are widely dispersed in UDC.

The principle of attaching applied science to fundamental concepts was advanced already by J. D. Brown. Now, after 60 years, it is fascinating to hear this voice from the grave:

> Its basis [Brown's system] is a recognition of the fact that every science and art springs from some definite source, and need not, therefore, be arbitrarily grouped in alphabetical, chronological or purely artificial divisions, because tradition or custom has apparently sanctioned such usage. The division seen in most classifications in vogue—Fine Arts, Useful Arts and Science, are examples of the arbitrary separation of closely related subjects, which in the past have become conventional, and it may seem heretical even at this late time to propose a more intimate union between exact and applied science.

Brown's system was perhaps imperfect in some ways, but it nevertheless seems remarkable that the statement above did not influence the international development in classification.

F. *Limitations on the Use of a Hierarchical Universal Document System*

The scheme need not enter more deeply into details than a generally applicable differentiation of documents allows. When we come down to a certain level, perhaps the 3- or 4-digit level, we can no longer find principles of classification which can be generally accepted. The best subdivision for one institute or library may not be suited to others. A certain institute may very well arrange a detailed classification for its own use, but in an international standard one should not go too deeply into detail. Even if a subdivision on modern principles into only a few groups could be generally accepted, very much would be gained. Other methods of search can then be employed both within a class and in combination with other classes. Faceted systems, coordinated indexing, or permutated indexing may have their place here.

The question whether one can classify a document in a document system so that it can easily be retrieved is tied up also with the size and the degree of specialization of the collection in which the document is to be stored. In a small collection comprising many fields of knowledge a single concept may suffice to provide a safe anchorage, and a universal document system is thus best suited for such a collection. Perhaps the widest application for universal document systems is in filing of documents in offices and for every man's use.

If the collection grows larger, still with unchanged composition, more documents will come under every class number. This can hardly imply that the possibility of finding relevant documents will decrease. But if we want to use this bigger collection with full effectiveness, it may be reasonable to try to refine the indexing and search methods.

If we now compare with a very specialized collection, it is quite clear that the advantage of the universality of the system becomes smaller in the same degree as the documents are assembled under fewer headings. To characterize every document, we are forced either to get the class numbers in question more detailed in a way that best corresponds to the collection or to use other methods of indexing.

We have here, of course, presumed that the aim of the system is to display the collection in the widest possible way, and not to concentrate documents according to field of activity, as the character of some groups of UDC seems to indicate.

G. *Different Applications of a Document System for Different Fields of Activity*

By fields of activity is meant fields with a common need for documentation, their own special journals, etc. These fields may be scientific or social, sectors of industry or professions. Every part of the universal scheme corresponds to one or more fields of activity.

We also have a hierarchy—apart from the hierarchy in the document system—that could be exemplified by:

All science	=	"universal field"
Technology	=	"suprafield"
Building and civil engineering	=	"field"
Architecture	=	"subfield"

With increasing specialization the special document collections are more important than the universal, and the first aim of a universal system is to be a tool for special collections.

If we try to apply the principles mentioned above and draw up system structures suitable for (1) the universal field, for (2) some suprafields, for (3) some fields, and for (4) some subfields of activity, it is not probable that we shall get the same result in all these cases. This is one reason why the system presented above (TUS) is adapted to a "suprafield" technology.

I do not think that it is suitable to aim at different document systems for the fields and subfields under technology, but it may be justified to have *different applications* of the same system. If we have a document "Electrical Installations in Buildings," the classification allotted when publishing the document will consist of two equivalent notations: *b* for building and *e* for electrical. The "electrical library" may use *b* and the building library *e* as primary classification. While this is a simple example, the problem as a whole is rather complicated and cannot be thoroughly studied without the background both of a certain document system and of the outline of a division into fields of activities and their documentation activities.

It is conceivable however that documents, as an aid both for filing in different collections and for distribution to the proper categories, might be classified both according to knowledge content and to receiver categories (fields of activity).

2.1.2. SPECIALIZED HIERARCHICAL SYSTEMS

Among the reasons why there are hundred or thousands of special systems in use in most countries, one is that the UDC numbers are too long and that specialized spheres of interest disappear in the universal collection. This is not a weighty reason, for one often overlooks the simple course of substituting the most heavily loaded numbers by letters and making a selection from the main system in which irrelevant titles are weeded out. A more weighty reason is that the structure of UDC has proved unsuitable for many particular fields. The absence of a main group for materials, for instance, seems to be a serious defect. Only a radical revision would suffice in face of such criticism.

If one could get all those who now use special systems to use a common universal system instead, the circle of users of the universal system would be multiplied many times over. The picture is, however, not complete without paying consideration to the faceted systems.

2.2. FACETED DOCUMENT SYSTEMS

The faceted systems drawn up in England are all intended for special fields. They are not, as are hierarchical specialized systems, constructed by dividing the special field of knowledge from above in certain parts. Instead they are constructed from *below* by organizing the most important concepts (represented by certain terms) in a limited number of series (facets) and the documents are characterized by means of one concept from some (or all) of the facets. Thus there is no precoordination of concepts involved. If we consider each facet as one classification system, the concepts represent the "subjects," but for the system taken as a whole the documents represent the subjects. The faceted principle is more related to indexing than to classification in general sense.

What is here of special interest is that members of CRG are striving at the design of a universal document system with faceted structure. Main facets are Things, Activities, and Properties. A series for things arranged according to "integrative level" have been published. For further information it is better to refer to various publications of the members of CRG, especially perhaps to Foskett's *Classification and Indexing in the Social Sciences* (14), where, moreover, a very interesting and exhaustive analysis of the whole classification problem is given. We quote from this book:

> What is certain is that a general classification scheme is needed, that none of the existing schemes are satisfactory, and that we have enough ideas about the structure of a new scheme to make the effort to produce one worthwhile.

If we compare this project with UDC or with the enterprise to design a new hierarchical document system along the lines advocated above or other projects going on, it would be better not to try to formulate an opinion concerning which may be the best way. Research and experiments in more directions and with organized comparisons may be better than any attempt to compel the development into a single track.

• 3. Other Systems Based on Knowledge Content

A. EDITORIAL SYSTEMS

Every author of a textbook, and every scientific writer, is faced with the same problems as the documentalist, the systemization of knowledge in some way or another; but in the editorial field these problems are not at all objects for organized studies. The editor of a technological compilation, for example, or of a code of standards, is presented with the problem of arranging a number of "facets." These must be chosen so as to bring out the most important concepts as headings; the number of facets must not be so small that they cannot accommodate the desired contents. Nor must it be too large, for then there will be too many entries. The editor will then have great problems in deciding on the choice of location of the knowledge elements and the reader will meet with difficulties in finding the information he requires.[3]

Just as important as it is to keep documents in order, so it is to bring order into their contents by editing the documents.

Here, as within documentation, we have the struggle between the alphabetical and the systematic principle. The alphabetical has a great significance in encyclopaedic works, especially the bigger ones of a general nature, and in books of reference, owing to the fact that the general public do not want to choose between different possible entries but go direct to a title-word they have in mind. Just as a document may occur several times in an alphabetical index, so information can be duplicated in an alphabetical work, perhaps linked together by cross references. For books intended for learning and education the systematic form is the obvious choice.

Editorial systems have significance on the universal plane as well. Compilations of this kind exist both with the classical disciplines (mathematics, mechanics, physics . . .) as headings, and with concept headings (number, space, time, mass . . .). Table 3 shows the systems in some modern encyclopaedias from different countries which conform closely to Table 2.

Books of this kind should follow the systematics of science, i.e., present knowledge in a natural sequence, implying that elementary phenomena are described before the more complicated and that new fundamental concepts are not introduced until the presentation so demands. Cross references should go backwards (toward the beginning of the book) and the need for them should be minimal.

B. EDUCATIONAL SYSTEMS

The most important phase in the distribution of knowledge is the education organized by society. Systems are necessary for dividing up the sphere of knowledge taught at each school and university into curricular subjects. Obviously these systems must be based on the system of science. It should be possible to give every subject a definition by reference to a universal system.

All educational planning should be based on a systematic analysis of the specified requirements of knowledge for different occupational groups. The relation between education and documentation is mutual and sig-

TABLE 3. Some Modern Encyclopaedic Works with Systematic Structure (by Subject)

Larousse Metodique (Librarie Larousse, Paris 1959)		
Aritmetique	Astronomie	Biologie
Algebre	Physique	Botanique
Geometrie	Chimie	Zoologie
Mecanique rationelle	Geologie	etc.
Universitas Litterarum (Walter de Gruyter, Berlin)		
Matematik	Mineralogie	Medizin
Physik	Palaentologi	Psykologie
Chemie	Botanik	Volkerkunde
Astronomie	Zoologie	Sociologi
Geologie	Anthropologie	etc.
Kleine Enzyklopadie (Verlag Enzyklopadie, Leipzig 1960)		
Natur		
Zahl	Kraft (energi)	Leben
Raum	Stoff	Tier
Zeit	Weltall	Pflanze
Mass u. Gewicht	Erde	Mensch
Fakta (Bokforlaget Fakta AB, Stockholm, 1955–1961) and *Facta* (Ediciones Rialp, Madrid, 1962–1966)		
Mathematics	Geology	Countries and peoples
Mechanics	Meteorology	Humanistic fields
Physics	Biology, general	Society
Chemistry	Plants, Animals	Technology
Astronomy	Man	Home and environment

[3] An application of a specialized system with decimal notation has been realised on a large scale in the Swedish building handbook BYGG, published in the 1940's in four volumes (in the 1960's increased to 6 volumes), with three numerals before the colon and a maximum of three after (e.g., 242:518 Concrete; strength; compression; testing). This system has also been used for filing of documents and is often preferred to existing library systems. The same notation principles but with a universal system are used in the general encyclopaedia Fakta (Table 3).

nificant, but have the relations between curricular subjects and document systems anywhere been investigated?

C. Systems for patents and legal texts

Both in a patent specification and in a code of laws, certain relations are established between concepts which have a legal force. The problem of classifying or indexing such texts is, as is well known, closely akin to the general problem of documentation and these questions have been dealt with in the documentation literature. In fact such texts should be ideal experimental subjects owing to their concentrated form and the care with which concepts are selected, clothed in word form and syntactically compiled.

• 4. Term Systems

A. General

Vocabularies in different languages have not been created at one time as the result of a general study, but have developed successively among different peoples insofar as they have felt a need for words. Science, technology, and social organizations are constantly producing many new concepts and require an extended and exactly defined vocabulary. The creation of new words is a slow process, and it is therefore not remarkable that the connection between concepts and words is poor. The nomenclature organizations have the difficult mission of guiding the development into the proper paths.

The stock of terms is of great interest in documentation. The knowledge contained in documents can, as is well known, be indexed, either by a number (or several numbers) in a document system ("item-entry") or by a number of concepts or terms characteristic of the document. If we index by terms ("term-entry") it is important to draw up a list of terms (descriptors or key words) which should be given priority both by the indexer and by the searcher. These terms are collected in a special word list, often called a "thesaurus."

B. Different ways of arranging terms

The commonest and simplest method of arranging terms is *alphabetically,* a convenient but, as we know, an often deceptive method. As complement to all other classifications, however, the alphabetical is indispensable.

Then we have the *grammatical* division according to parts of speech which determine to a certain extent what syntactical role the word can play. The trilogy things, activities (processes), and properties correspond by and large to the three main parts of speech, the substantives, the verbs, and the adjectives, but the context between concepts and parts of speech is, as well known, not always ambiguous. Prepositions and other small words explain—together with a more or less extensive system of case-declensions—the syntactical context between the words. The logic of natural speech, however, seems to be insufficient in qualified information retrieval, and different sets of "role indicators" have been devised by different documentation schools or experts.

There is also the *etymological* classification, which is based on the origin of words and cuts across the grammatical classification. Words are organisms which are born, develop, change their meaning and spelling, and sometimes die. The etymological classification corresponds to the natural genus-species division in the plant and animal kingdoms. The synonyms are representatives of different etymological individuals, joined in the same concept category. The words "hound" and "dog," for example, come in different etymological places but in the same concept category and so do "hen" and "poultry." Composite words correspond to biological crossings.

We can also classify words *semantically,* on the basis of the concepts the words are to convey.

Concepts can to some extent be expressed by *symbols* or *pictures* without the help of words. Geometrical figures can be easily illustrated; the elements, too, by means of their electron shells; chemical compounds by their molecular diagrams; cells, tissues, plants, and animals can be illustrated, as can technically produced objects. (Maps, diagrams, and tables can also exhibit very complicated relations without words.) Illustrated glossaries are not unknown and have a function to fulfill within documentation.

The language of symbols is an intermediate step between concepts and words. Can one imagine a complete universal system of systematically arranged symbols?

Terms can also be expressed by *definitions,* a new term being defined by means of other already defined terms. This implies a certain systematic arrangement, e.g.:

3.	A figure bounded by *three* straight lines:	triangle
4.	A figure bounded by *four* straight lines:	quadrangle
4.1	Ditto with right angles:	rectangle
4.11	Ditto with right angles and equal sides:	square

Here, as a matter of fact, we are approaching the editorial principle of presenting knowledge in text books by building a system of concepts step by step (cf. Section 3).

To a certain extent terms can be easily arranged in groups which are, more or less, generally accepted, such as the elements, chemical compounds, rock species, plants, animals, etc. If we strive at a fixed unambiguous location for every word, we must always keep to the fundamental meanings of the words. Granite, for instance, is a geological term (a species of rock) and not a building material or a product from the quarrying industry, for the definition of the word belongs to the sphere of geology. Rose is a botanical term; dog, a zoological.

Lists of terms arranged with regard to the semantic signification of the words are needed within documentation. A scheme for a complete, systematic list of terms, in which each term is as far as possible located according to its meaning, we shall here call an *unambiguous semantic term-concept system* or, more simply, a *term system.*

The elementary terms in the sciences are generally not created for the sciences, but for more primitive needs of man. Hot and cold are originally not physical terms, they are expressions for certain feelings of man. However, the sciences have need of such elementary terms, and allot them fixed positions and clear definitions. We will get a good guidance for an unambiguous location if we give preference to the more fundamental locations in the series of the sciences and concepts as shown in Table 2.

To a very large extent even terms considered to be "general" can thus be assigned a place according to the sciences. Large, small, increase, decrease will go to Mathematics; long, short, round to Geometry; rapid, slow to Kinematics; force, equilibrium to Statics; hot, cold to Thermodvnamics; beam, reflex to Optics; granite and syenite to Geology, etc.

C. SURVEY OF EXISTING TERM LISTS

One might have expected there to exist a systematically arranged glossary of all our technical terms; at all events in some language; in fact there appears to be none, at least none that is up-to-date.

A well-known, systematic glossary of universal character is Roget's *Thesaurus* published in England in 1852—which was enlarged in 1952 (15) and has a wide use as dictionary of synonyms for writers and editors. The original system has been retained, but this is unacceptable for modern science and technology as physical and technical concepts are derived from the human senses, which leads to corresponding gaps in physics and technology. Example:

- Matter
  - Organic matter
    - Sensation
      - Touch, Heat, Taste

- Intellect
  - Communication of ideas
    - Modes of communication
      - Publication
        - . . . journalism,
        - . . . radio,
        - . . . television

Coordinate indexing has brought a new wave of interest in technical terms, and attempts are being made within all branches of technology to perfect their stocks of words. There have been discussions concerning the use of UDC as a systematic term list; but UDC is not constructed as a term system, and the same terms may occur 5 to 10 times or more, whereas a large number of the elementary words are missing. One would first have to rebuild UDC so that the same term occurs only once, and the principles of such adjustment of the system would then have to be investigated, which to some extent is equivalent to constructing a new system.

It is imaginable, however, that large parts of UDC, e.g., within chemistry and electrotechnics, could be used. And as UDC is published in several languages, the use as far as possible of this system, and possibly other classification systems, should be valuable.

Example from the trilingual edition of UDC:

German	English	French
621.375		
Rohrenverstarker	Valve amplifiers	Amplificateurs a lampes
Magnetische Verstarker	Magnetic amplifiers (transductor)	Amplificateurs magnetiques
Kristallverstarker Transistoren	Crystal amplifiers (transistor)	Amplificateurs a cristal Transistors
Dielektrische Verstarker	Dielectric amplifiers	Amplificateurs dielectriques

A glossary of terms common to the whole of technology and intended for coordinative indexing is the Engineering Joint Council (EJC) *Thesaurus* (16). The terms are alphabetically arranged, but with references to "broader terms," "narrower terms," and to otherwise related words ("refer to"), and from words which should not be used as descriptors to the respective descriptors. These arrangements may be interpreted as a system of "microhierarchies," but no attempt has been made to bring together all terms into a single systematic totality.

A number of special thesauri have been compiled for different fields, and in them one sometimes finds a systematic classification of the terms of that field. In a thesaurus published by Euratom (17) for nuclear technology, for instance, we find a classification of the vocabulary into 42 groups (key word groups). It may be of interest to see whether two institutions select and arrange the terms within the same field in the same way. Within the field of radiation and electromagnetic waves Euratom arranges the key words in four groups:

72. Optics
73. Magnetism
84. Radiations
85. Elementary particles

If we compare the 17 key words under Radiation in Euratom with the 28 "narrower terms" under Radiation in EJC, we find that *only one* is exactly the same in the two groups, namely Cosmic Radiation, while the following three are similar in their import but differ in spelling and linguistic form.

Cerenkov radiation	—	Cherenkov radiation
Gamma rays	—	Gamma radiation
X-rays	—	X-radiation

Otherwise the selection of terms is entirely different in the two thesauri. The compatibility between them is obviously very poor. As regards special thesauri the following points may also be noted: The barrier between science and technology that is characteristic of UDC does not exist here. The key word groups thereby have a more homogenous character, and the structure is more rational.

Concerning terminology, on the other hand, one notes a diverging tendency since each special branch of technology uses general terms in a specialized sense for the branch. In a thesaurus for roadbuilding, words such as base, bed, carpet, coat, deep, junction, etc., appear in the key word group "Road." This may be acceptable if road documentation is to be considered an isolated activity; but if this activity is to be included in a broader context, one will have to deal with different vocabularies. Closer agreement between terms and headings is found in key word groups such as "Mathematics—Mechanics" and "Optics."

An interesting example of a list of terms is *Medical Subject Headings* (18) of *Index Medicus,* a list of biomedical terms arranged both alphabetically and systematically ("categorized list"). We quote:

> The Categorized List is an attempt to bring related terms together where the indexer and searcher can view a panorama of subject headings and select the most appropriate heading for his needs.

The introductory categories are:

A. Anatomical terms
B. Organisms
C. Diseases
D. Chemicals and Drugs
E. Analytical, Diagnostic, and Therapeutic Technics and Equipment

Example of subcategories:

A1. Parts of the Body
A2. Musculoskeletal system
A3. Digestive system

The terms ("subject headings") have usually been unambiguously locatable within a main category, but thereunder they often occur within several subcategories. Example: "Hand" occurs both under A1 and A2. As cross references are given, this perhaps is of no particular inconvenience, but by adjustment of the category headings it should be possible to achieve unambiguous locations to a greater extent. Generally speaking, one may say that A, B, and C of *Index Medicus* could be incorporated in a universal system of terms.

It is interesting to compare these categories with the headings which are usual in document systems:

Index Medicus (term categories)	*UDC* (subject fields)
A. Anatomical terms	58. Botany
B. Organisms	59. Zoology
C. Diseases	61. Medical Sciences

The pattern in the "realm of knowledge" is not the same as in the "realm of terms." Botany, Zoology, and Medicine have largely the same stock of terms as the groups A, B, and C. But if terms are collected and arranged from the bottom *upwards* as in a thesaurus, one comes automatically upon certain groups which do not always agree with the traditional sciences and are found in document systems for which the classification is made from the top *downwards.*

One can easily find out which classification of a collected stock of biomedical terms results in the largest number of duplicate locations. That, at all events, Zoology and Medicine have a common stock of anatomical terms is natural since, anatomically, man is an animal. Thus the A, B, and C grouping above must in this case be better for the term system. But the following term categories are also conceivable:

Botany	Zoology and Medicine
Anatomical terms	Anatomical terms
Organisms	Organisms
Diseases	Diseases

In a universal term system, of course, the group Chemicals must not come under the heading Biomedicine, but appear earlier in the systematic sequence. One may then perhaps find that Drugs as well can be removed from the medical field.

In the field of materials technology I have made some attempts to systematize the stock of terms (materials, processes, properties), and the result suggests that an unambiguous systematic grouping of these terms is not a hopeless task. It is also quite clear that one arrives at hierarchies of an entirely different kind than in a document system, as will be shown.

In the section on document sytems (2.1.1.C) we have given an example of a possible subdivision of a document collection on concrete, corresponding to the composition of the stock of documents. In a term system only words relating exclusively to the concept "concrete" may appear under the heading—words such as standard concrete, lightweight concrete, etc.—while terms for properties and processes appear elsewhere. In the document system the hierarchies are to a certain extent expressions of combinations of concepts which have been regarded as suited to a collection of documents and a document on "Strength of Concrete" has here a fixed location. In the term system one is concerned only with generic hierarchies (each term being defined by superior headings). The same document has to be indexed with help of two terms (as in a faceted system). (Fig. 2.)

A universal, systematically (and alphabetically) arranged thesaurus can be created by successively sifting the vocabularies from different subject fields through the "screen" of a universal system. The dilemma is that a universal term system is lacking. There are two ways of creating such a system: by way of theoretical speculation or, practically, by proceeding from a universal document system. This must, however, be successively adjusted to a systematized arrangement adapted to terms,

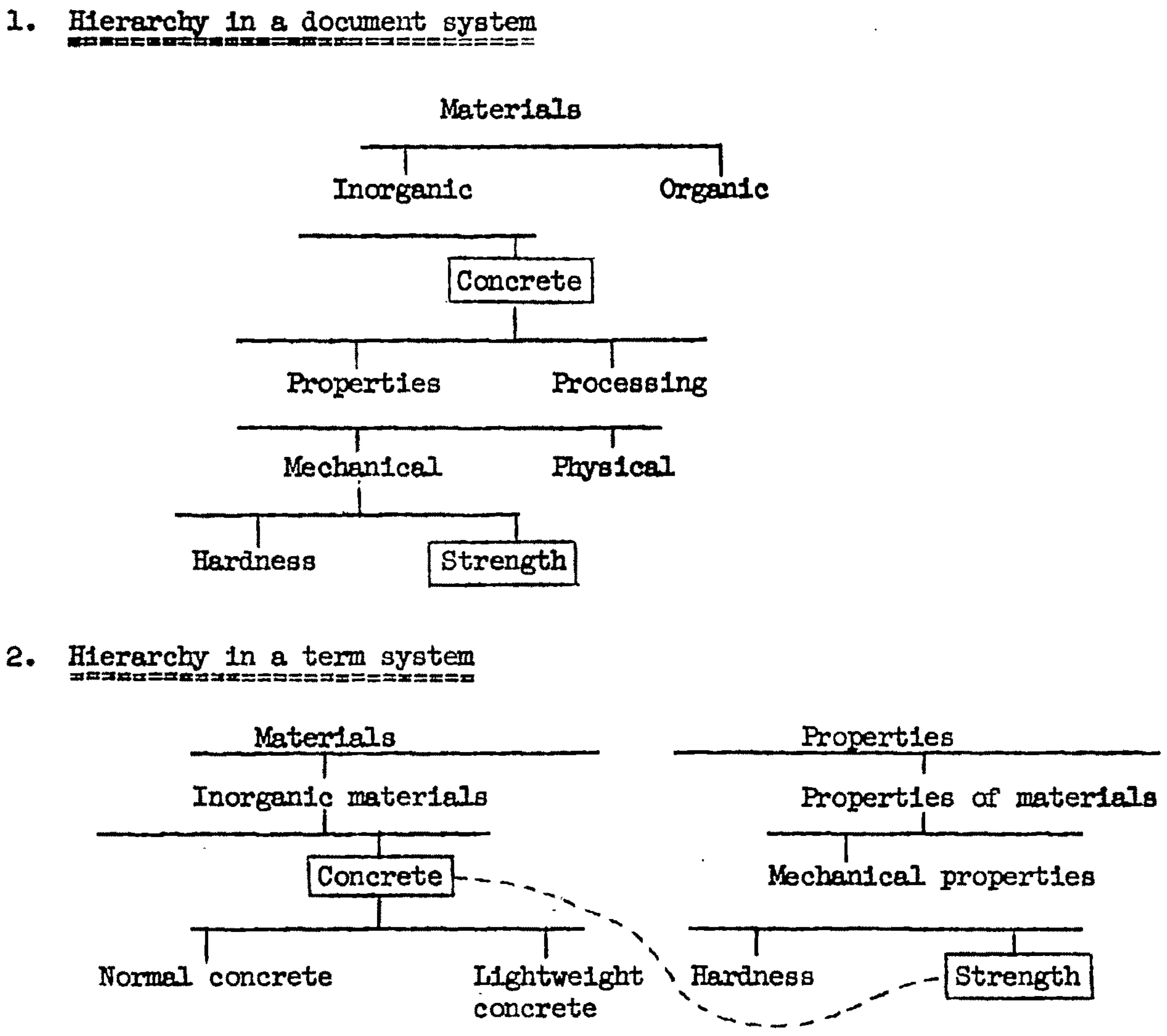


FIG. 2.

as is shown by the comparison between *Index Medicus* and UDC. The more conceptual and the less bound to tradition the system used, the easier it seems to be to transform it to a term system.

Attempts on a smaller scale with assortments of vocabularies from different quarters have been started with the help of TUS as a provisory term system. Every term is given a number according to TUS and then sorted both on the basis of this number and alphabetically. With the aid of punched cards a project of this kind can be very well carried out on a large scale and the value would be greatest if the work could be conducted on a universal basis.

To avoid misunderstandings it should be emphasized that a term system constructed on the principles presented here will be quite different from a concept system for classifying documents. It has been shown that a document system of the normal model is not suited as a term system. But can a term system be used as a document system or as the framework—in one sense or another—of a document system?

Obviously many traditional subject fields will be broken down (cf. *Index Medicus*). But if we think not of the universal libraries but primarily of the special libraries where often this breakdown has already been made, it may be wise to postpone expressing an opinion until the question has been examined.

Opinions among documentation experts concerning the possibility of creating a universal system of terms can be illustrated by some quotations. Vickery (19) puts the question:

> Is a universal descriptor language possible which embraces all subject fields, and is usable in all retrieval systems? . . . Perhaps the present area of specialization in retrieval will be succeeded by a synthesis, leading to the general use of a common "interlingua."

This question comes back unaltered in the second issue of his book (1964). Evidently nothing has happened in three years that has made it possible for Vickery to give an answer to his question.

Mortimer Taube had some doubts on the possibility and wrote: ". . . no one has actually produced a general categorization or classification of a total vocabulary."

The universal facets of CRG must in this connection be of interest as Foskett presents the theory of integrative levels as a principle for dividing terms into groups. The relation between document systems and term systems here comes in a new light.

• 5. Practical Systems

By practical systems is here meant systems used in industrial enterprises, in trade, and in other practical connections, aimed at the classification of products, costs, etc., and for filing purposes.

A. PRODUCT SYSTEMS

Systems for technically produced things we may call *product systems.* All industrial enterprises need such systems to maintain order in their manufactures and often spend large amounts of money on them. They are used for stockkeeping, for registration of drawings, for filing of documents on products, for costing and industrial planning. Often an entire trade has a common system (e.g., the hardware trade, the furniture trade).

Products can be classified from different points of view. They cannot as natural things (plants, etc.) be classified according to genus-species relations, but the raw material and the industrial production process offer a counterpart to some extent. Material, shape, and purpose are for many products the most important factors; and a concrete drainpipe may thus be classified as *pipe,* as *concrete product,* or as *products for sewage.*

These remarks apply primarily to bulk materials and to semimanufactured products. The classification of *machinery* and *apparatus* offers greater difficulties since, in this case, the principle of function (a particular physical principle, for example), or a category of user, may offer the best classification (e.g., electrical or optical equipment, office supplies, etc.). A universal system would seem to be necessary as a basis or background for compiling a comprehensive system for the latter categories.

The degree of complexity can also be used as a first principle of classification, and under the main groups derived in this way various principles of subdivision can be employed.

Product systems in industry are often used also for classifying documents concerning products; this is sometimes their main aim. Sometimes these systems acquire the character of complete document systems, with space also for sciences, social questions, and the like. In such case they often compete with the system used in the library, which in turn may be elaborated in detail into a product system. UDC tends in this direction but is generally not used as product system; as a rule industrial enterprises draw up their own systems which they consider better suited to their specialties.

The existence within the same firm of two or more schemes with different structure, but comprising partly the same concepts and terms, cannot be an ideal solution. This now seems to be the rule, however, and a big firm or institution sometimes has even three, four, or more schemes which perhaps could be replaced by a single one.

The need for a document system which is also a complete industrial product system—or at all events constitutes a basis for such a system—is fully legitimate. One may then ask whether this is an unreasonable requirement, or can one imagine a universal classification system which can serve satisfactorily both as document system and as a basis for a universal thing-and-product system?

It is naturally not possible to construct a universal product system simply by adding together a number of such industrial systems, as they usually overlap to a great extent. Perhaps a universal product system could be made as a faceted system with facets for material, shape, function, etc.

A standardized facet for materials and another for form may be the best way to start with, if we strive at some uniformity in this field, and these facets could in some way be included in a universal document system. The function facet could also be a connecting link with a universal document system.

B. SYSTEMS FOR PROCESSES AND LABOR

Materials and products are the "substantive facet," in industry, processes, and labor the "verb facet," and properties the "adjective facet." Processes include both intellectual work and manual and mechanical work. In this context we shall not be concerned with intellectual work. To get a classification for processes and labor is more difficult than for materials and products.

Some processes can easily be classified according to processed material (in the building trade woodwork, concrete work, etc.), but not all material processes or material treament can be classified thus, as they are applicable to all or some kinds of materials (e.g., welding, drilling). Other processes must be classified on the basis of the finished product (e.g., installation or assembly work) or as "auxiliary processes" which cannot be related to either a particular material or a product. However, it is not impossible to get a generally acceptable division of the processes in industry; such a division is essential and has already been accomplished thousands of times. We have in fact only to consider which solution is the best one, if we want to get a standardized system of this kind.

It would be possible to obtain some correspondence between systems for processes and systems for materials and products, and it also seems possible to include a category of material treatment processes in a universal document system.

C. SYSTEMS FOR PROPERTIES

It seems to be difficult to construct a universal system for properties of all kinds. For the properties of materials, on the other hand, a system associated with the systematics of science may be fairly easily constructed, as certain special institutes have devoted much effort to studies of all the properties of materials. The study of materials indeed consists to a large extent of determining their properties expressed in units of measure. These units, which are composed of a limited set of basic units (cf. Section 1), are a good guide to the systematics of properties, which is closely related to the classical subdivisions of the natural sciences:

Mechanical properties
Thermal properties
Electrical and magnetic properties
Chemical properties,
etc.

A system for properties of materials is valuable for all industrial enterprises producing materials in their specification of properties. It is also of interest for testing materials and for general consumer information. This is a necessary step also when drawing up a list of terms for technology.

Summary

The purpose of this paper is to point out the fact that we in documentation have use not only of one kind of classification systems but of different types distinguished by *different types of subjects*.

The discussions are often confused as the primary purpose of the classification is not made clear. Some experts talk of classifying knowledge, others of classifying concepts, ideas, terms, etc. We will get a more fruitful and pragmatic approach if we consider the type of subject of the general library classification systems to be *documents*, even if the system is constructed on the basis of fields of knowledge or on concepts.

In documentation we also have a need of a classification system for terms, a *term system*, based on the semantic signification of the terms. The demands on the hierarchical structure is always determined by the subject thus for term systems by the terms themselves. The patterns in the "realm of knowledge" is not the same as in the "realm of terms." To investigate principal differences in these different patterns seems to be an important task.

Further on we have in all industrial enterprises, etc., need for practical systems with different objects, e.g., *systems for products*. These can be used for specifications of products, for cost calculations, etc., and also for filing of documents on the products.

It is important in developing classification systems of different kinds to strictly have the subjects in mind. It may be that we want to use a document system for classifying terms or a term system or a product system for classifying documents or a document system for classifying products, but we have to be aware of the fact that we then are using these systems for a secondary purpose.

References

1. BRUGGHEN, W. VAN DER, Report on the FID in 1964, *Revue Internationale de la Documentation,* 32 (No. 2): 60, 1965.
2. BLISS, H. E., *The Organization of Knowledge and the System of the Sciences.* New York, 1934.
3. VANNÉRUS, A., *Vetenskapssystematik,* Alb. Bonniers förlag, Stockholm, 1921.
4. BACON, FRANCIS, *Of the Proficience and Advancement of Learning,* 1605.
5. DE GROLIER, E., *A Study of General Categories Applicable to Classification and Coding in Documentation,* Printed in the Netherlands, Unesco, 1962.
6. DOBROWOLSKY, Z. (Warzawa), *Analysis of Classification Systems,* Report to the 2nd FID/CR Meeting in Stockholm, 1963. Stenciled copy.
7. BROWN, J. D., *Subject Classification,* London, 1906.
8. TULLANDER, B., Vetenskap med och utan metod (Science with and without method), *Svenska Dagbladet,* Stockholm, March 1965.
9. KORACH, M., The Science of Industry, Chapter 13, *in* N. Goldsmith and A. McKay, eds., *The Science of the Sciences,* Bungay, Suffolk, 1964.
10. *Universal Decimal Classification,* Trilingual abridged edition, British Standards Institution, London, 1958.
11. WÅHLIN, E., Principles for a universal system of classification based on certain fundamental concepts and an outline of a variant adapted to technology, *Journal of Documentation,* 1963 (4):173–186.
12. WÅHLIN, E., *Basic Principles and Outline of a New Universal System of Classification,* Proceedings from the Conference on Building Documentation, Genève, 1949. Stenciled copy.
13. WÅHLIN, E., *25-year Bibliography for the Journals Byggmästaren and Arkitektur,* Byggmästarens förlag, Stockholm, 1966.

14. FOSKETT, D., *Classification and Indexing in the Social Sciences,* Butterworths, London, 1963.

15. *Roget's International Thesaurus,* Thomas Y. Cromwell Co., London, 1962.

16. *Thesaurus of Engineering Terms,* Engineering Joint Council, New York, May 1964.

17. *Euratom-Thesaurus,* Presses Academiques Européennes. Bryssel, February 1964.

18. *Medical Subject Headings,* National Library of Medicine, 1965.

19. VICKERY, B. C., *On Retrieval System Theory,* Butterworths, London, 1961.

Letters to the Editor

June 23, 1966

Dear Sir:

Ashley Speakman reporting (*AD Newsletter*, May–June, 1966, pp. 2–3) a panel discussion on "Intellectual versus Mechanical Indexing" (at the May 12–13 "Colloquium" on Information Retrieval) writes the following:

> . . . intellectual indexing is the order of the day, primarily for economic reasons; and mechanical indexing will replace it, particularly in large information systems. There was agreement on this, but not on when.

Mr. Holm and Mr. Liston (panelists) can speak for themselves, if they care to, but it was not my impression that they agreed with this thesis. In any case a number of statements and questions from the audience (more than an hour was devoted to audience-panel interchanges) indicated that the meeting as a whole did not.

Was the *Newsletter* report written by Mr. Speakman or by an experimental computer program?

John O'Connor
R.D. 1
New Tripoli, Pennsylvania

July 7, 1966

Sir:

During the past seven years as Director of the Office of Documentation at the National Academy of Sciences-National Research Council, it was a part of my responsibility to be familiar with science information projects in many subject fields. Among these were (in no particular order):

Behavioral sciences
Brain science
Textile and apparel research
Highway research
Food and Nutrition
Automatic language processing
Chemical information and coding
Critical data
Prevention of deterioration
Geological sciences
Building research
Biomedical communication
Cardiovascular literature
Nuclear physics
Foreign science
Metallurgy
Pacific science

The point I wish to make is the *similarity* from a professional documentation point of view of the problems among this wide spectrum of subjects. Over and over again the fundamental problems of procurement, languages, classification, indexing, etc., came up. It is more than ever my firm conviction that a *real subject* resides in science information work, and I feel I would be remiss if I did not express this as forcefully as I can on the basis of my experience. I wish to do this now especially, as others have questioned the validity of a documentalist's approach.

Even a casual inspection of the above list will show some apparent overlap, but this makes my case even stronger, as the initial viewpoints were often quite different in apparently related areas.

Karl F. Heumann
6410 Earlham Drive
Bethesda, Maryland

July 7, 1966

Sir:

In addition to my congratulations on *Documentation Abstracts* which I have expressed elsewhere, may I make two points as a result of careful reading of the first issue:

1. The section *Classification* contains a total of 22 entries, divided into 5 original articles from the U. S. and 17 from other countries. In my opinion this is a reasonably accurate reflection of the way the interest in this subject is divided.
2. The index parameters chosen ("Author" and "Anonymous Titles") are good first choices, but I want to suggest one other easy one. I refer to the possibility of using the many acronyms as entries. I have found 23 such items from AESOP (262) to UPLIFT (266), comprising 30 total entries, and I submit such a compilation would be both simple and useful.

Karl F. Heumann
6410 Earlham Drive
Bethesda, Maryland

Dear Sir:

Most discussions of the information retrieval problem start out by stating that information is a national resource. I wonder if we might not get a better perspective of our problem if we were to regard information as a waste-product.

Considering information to be a resource carries the idea that this pile of paper and words has great intrinsic value, that the quantity of it is limited, that its use must, somehow, be budgeted lest the source run dry. It seems to me that this corresponds very poorly to the true situation.

In fact, information is a by-product of most of our scientific and technical activities. There is so much of it that our capacity for disposing of it is overtaxed. The scientist writes papers describing what he has done. The engineer writes reports. The inventor prepares patent disclosures. All of these are by-products. Some have no value at all beyond the initial distribution, recording of receipt, and a single reading by the receiver. The document describing even the most important discovery has only a limited useful life. If it has great significance, it will soon be boiled down as part of a book covering the area to which it relates. If its significance is minor, it will be noted by the author's peer group and then neglected. In a good many cases, the information will have been passed on to the peer group in oral or preprint form well in advance of its publication. The published material will have its primary value as a source document for a later review or a still later book.

A document, once printed, tends to remain in existence. Perhaps even more important, we tend to regard every document as something that must be abstracted, indexed, SDI'd, KWIC'd, microfilmed, mag-taped, and on and on.

Isn't there a good possibility that most of those documents have lost almost all of their value by the time they get to the abstracting service? Isn't there a good possibility that the pile of paper is really a slag heap, not a mountain of pay dirt?

HARRY BAUM, *Director*
Technical Meetings Information Service
New Hartford, N. Y.

May 11, 1966

Dear Sir:

For some time it has concerned me that no one has written the history of ADI to date or has taken effective steps to gather the materials necessary to such a history. Recently some of my students struggled to write papers on the subject, and I remember that Dr. Kaiser, former Executive Secretary of ADI, had great difficulty assembling a complete set of AD, compiling a list of past presidents, and so on. These activities have pointed up the need for better housekeeping, but I think they point to something more. We have Dr. Watson Davis, founder of ADI, Dr. Luther Evans and Mr. Scott Adams, both past presidents, and many other interested and knowledgeable people within the New York-Washington area as first-hand sources of unrecorded information. We should be motivated to confer with them, use what they can give us to fill in the record and write a more interesting and authoritative history than one which, in the future, will have to be based on secondary sources alone.

I am about to present a proposal which could have gone directly to Council, but I thought there might be value in making it public via this letter so interested persons could send their reactions and, perhaps, offer suggestions, information or materials that bear on the proposal.

It is proposed that the Council of ADI appropriate $750 for compilation of an official history of ADI, to be published in American Documentation as soon as it is completed; $500 is to be an honorarium for the compiler, the remainder, a travel fund for the compiler to interview the persons judged best potential sources of useful information. Cost of gathering or reproducing pertinent source materials is to be charged separately to ADI; these materials are to be deposited at the close of the project for incorporation in the official ADI records. This proposal is to be altered or enlarged upon at the discretion of Council when it comes before them for action.

One of the graduate students at Drexel Institute has written a partial history during this school year that is most readable, but its content is limited to information from secondary sources. She, Mrs. Lillian Shreve, has expressed interest in being considered a candidate for the job of compiler should the Council act favorably on this proposal in the near future.

CLAIRE K. SCHULTZ
Senior Research Assoc.
IAMC
Line Lexington, Pennsylvania

Book Reviews

10/66–1R **Faceted Classification Schemes.** 1966. Brian C. Vickery. Rutgers Series on Systems for the Intellectual Organization of Information, vol. 5. Graduate School of Library Service, Rutgers University, New Brunswick, N. J. 108 pp.

It is not, I suppose, the place one would expect to find it, but volume 5 of the Rutgers seminars has for the reader of *Othello* and *The End of the Affair* a similar treat in store: a constantly problematic time-scheme. Those who have read volume 2 in the same series (J.-C. Gardin's *SYNTOL*) have had the same sort of tangled skein offered to them already. If it does not seem to me too much to expect that the presentation itself be simply reported, followed by the panel-discussion comments—except that the size of such a transcript might demand some abbreviation.

This has not been done here, however. Each of the Rutgers seminars is intended to present one system (or, as here, one family of systems), but to do so not in a textbookish style; the lecture style plus discussion is to be the instrument of this opening up of the texture of the argument. But if the whole thing, instead of being edited (with, perhaps, substantial abridgement), is given back to the author of the central lecture, who may then turn it into a collage displaying (as here, and as with Gardin's) a totally different seriality than the event itself constituted, we may well expect some degree of confusion. The confusion arises when the central lecturer, forgetting that his readers cannot, like him, relate the new string of (printed) words to the original string of (spoken) words, proceeds to explain things to himself while mystifying his readers.

For this reason, perhaps, but possibly for others as well, the Gardin volume (though most of its reviewers have been hesitant to tick it off properly) was a near-total failure to convey what it intended to.

Or at least it seems to me that there is some intention to convey something: Mills' volume on the UDC and Ranganathan's on the CC convey something. But neither Gardin nor Vickery seem to be quite sure (*a*) of what it is that they are supposed to convey, or (*b*) of what sort of audience they are aiming at. In the volume at hand, for instance, there are extended passages in the early pages (but, presumably, chronologically posterior to the rest of the volume), drawn from remarks of rapporteurs such as Rees and Mooers, which are about classification in the most general sort of way possible. Now I will admit that it is wise to lay such a groundwork of generality whenever one intends to launch out into unconventional theorizing, or into criticism of some particular manifestation of such general principles; but when the sequel is nothing but a comment upon a description of a single family of classifications, and when the necessary background is simply the central lecture itself, the conclusion cannot but be that the work as a whole is intended to convey introductory-level information. But if this is the case, the central lecture becomes all the more crucial in its original unity: discussion, especially if it tends to be corrosive, is appropriate to the report of a "seminar" intended for a reading audience thoroughly aware of the fundamental principles and their manifestations; statement (though not, hopefully, dogmatic; nor even, necessarily, apodictic) is the appropriate vehicle for introduction.

It would be well for an introduction to the faceted family of classifications to be available to the large numbers of American classifiers unfamiliar with them, and therefore unaware of the advantages of the analytico-synthetic approach as the basis of mature criticism of the current strategizations most in use by them: Library of Congress Classification (and Subject-Headings), and Dewey Classification. Such critical evaluation can best (or at least most easily) be made in terms of comparisons of the epiphenomena of the underlying principles, rather than of the principles themselves, as stated in such theoretical form as (say) in Ranganathan's *Prolegomena to Library Classification* (London, Library Association [2], 1957). Once the results are seen, these more difficult backgrounds can be explored. However, and here we come to the central failure of the volume in hand, it is not with such an introduction that we seem to be dealing.

Sections I and II ("Introduction to Faceted Classification" and "Aspects of Information Retrieval") are rudimentary, to say the least. And yet a good many passages are technical, not to mention highly speculative. Such speculation is of course to be encouraged if it leads somewhere—even in an introductory work. But the statement (p. 46) that "by combining terms in compound subjects [faceted classification as a technique of conceptual bibliography] introduces new logical relations between them, thus better reflecting the complexities of knowledge" in no way justifies itself; nor does it lead anywhere valid, since while it is indeed true that concepts not previously juxtaposed may well demand a new logical relation (or we could say "*imply* a new logical relation"), this relation is by no means given to us in any sort of helpful way by the fact of the juxtaposition. Again (p. 57) we find that "the only argument in favor of the use of symbols in a classified catalog is the actual classified arrangement of materials," for the goal of generic/specific strategization of search. This is true, but it is not the *only* argument, since without the symbols attached to the terms there can obviously be no such thing as a class-schedule at all—unless we are willing to go all the way back to an alphabetico-classed arrangement. Hospitality, next (p. 57), is erroneously defined as the ability of notations to be "free to extend"—which misses the crucial point about hospitality (either in array or in chain) that it is what allows for intercalation of new notations for new concepts. On the next page "intercalation" is used in the sense of the drawing in of notations from a more general classification, such as the fairly common use of the UDC (1/9)-table for place as penumbral to many special schemes. Finally, many librarians (and bibliographers, too, obviously) will take exception to the too-common documentalist-opinion (p. 70) that bibliographical description is "basically clerical."

In a broader criticism of the book (as well as of several other members of the Series), it must be asked why there is no index; what do discussions of searching methods tell us about the (family of) system(s), unless they are methods peculiar to the system itself; and, most fundamental of all, how can one system be adequately described without comparison to others? Thus, not all the blame for the failure of this volume can be placed with its author: his handicaps are many, and the temptation to try to weld the pieces together after the fact is not really an advantage. Yet Vickery (and Aitchison, in her "Case History: The English Electric Scheme" [section VIII]) has managed to outline the *process* by which a faceted classification is prepared and used. This is done sometimes not in the order most helpful to the uninitiate, though each difficulty is eventually explained; it is done without overconfidence in the salvific power of such techniques, but at least sanguinely; but it is done, I fear, without adequate theoretical foundation. What we have here is a torso, perhaps helpful, but characteristic of the series to which it belongs; one can only hope that it whets the appetite of those who need it, even if it does not totally satisfy it.

JEAN M. PERREAULT
Lecturer
School of Library and Information Services
University of Maryland
College Park

10/66–2R **Computer Typesetting: Experiments and Prospects.** 1965. Michael T. Barnett. MIT Press, Cambridge, Mass. 245 pp.

While Director of the Cooperative Computing Laboratory (CCL) and Associate Professor of Physics at M.I.T., Dr. Michael Barnett carried on developmental work in the field of computer typesetting. After first seeing a Photon photo-typesetter in operation in a commercial plant early in 1961 he began a project that summer to activate that machine by means of computer-produced paper tape. A grant enabled him to enlarge his staff and increase his efforts. The work he describes in his book continued until early in 1964. At that time he returned to England and completed this study while at the University of London. Now back in the United States, Dr. Barnett is Staff Engineer in the RCA Systems Division at Princeton, N. J., having apparently been bitten in earnest by the typesetting bug.

Barnett's book begins with a description of the computer programs written for typesetting purposes at CCL and an account of the manner in which the Photon device operates. At the outset of the project he knew nothing about the intricacies of composition, either in "hot metal" or on film, but he learned quickly and, as a pioneer in this field of computerization, achieved some significant results.

He was not the first to conduct experiments of this nature, utilizing computers, but at that time he was ignorant of the work that others before him had performed. He was not aware of the comprehensive study, known as the Rome Air Force Project, subsequently published (in 1962) at the Thomas J. Watson Research Center* or of the experiments with Linofilm composition for final output in Russian-English translations. Nor was he familiar with the efforts expended in the newspaper field where programs of "justification and hyphenation" for "hot metal" composition were currently under development with the support of such major hardware suppliers as IBM, GE, and RCA. And yet, without the practical assistance he could have derived from broader exposure to the requirements of the printing and publishing industry, he and his staff created some very respectable programs which went a long way to arouse interest, especially in the scientific community.

Barnett was inevitably plagued with hardware incompatibilities and his method of producing input to the computer was cumbersome: eight channel Flexowriter tape to punched cards by way of a modified IBM 047 paper-tape-to-punched-card converter and ultimately into an IBM 32 K 709 computer. The magnetic tape output from the 709 was then read at the M.I.T. Lincoln Laboratory by the IBM 1401 and 1012 attachment to punch eight channel paper tape, two frames of which are required to describe one character for Photon 560 input. In the early chapters of the book he describes the equipment employed and certain of the features of the software packages he and his staff created. Too much detail is included to hold the interest of the casual reader; not enough is provided to satisfy the curiosity of one who wishes to become intimately familiar with the problems and solutions encountered. One of the major but unavoidable shortcomings of current publications in the field is the difficulty of defining the characteristics of the readership circle, which may or may not include computer people, printers and typesetters, publishers, librarians, information-handling specialists, economists, and automation sociologists.

The interest of this reviewer was to discover whether Barnett had succeeded in identifying the same problems we at Rocappi had encountered, both with respect to software design and more especially in the vital area of man-machine relationships. It is not so much in discovering ways to make typesetting machines perform that the challenge exists. It is rather the development of an operating system which takes into account all of the diverse elements that must be brought together for the successful production of a book. Solutions which evolved under Barnett's direction are much like those which we employ and, like him, we came to our techniques more or less in ignorance of what others may have developed. One looks for parameters which provide useful "hard copy" counterparts and have some meaning in the trade. One develops format shortcuts and one begins, little by little, to explore the data-processing applications wherein typesetting emerges as a by-product or an incidental but useful end result of other computer processing activity, or conversely, the data-processing by-products which may be derived from typesetting.

Barnett's group developed two basic types of programs: one accepted non-fielded input of the "sentential" variety and justified it without hyphenation by expanding the interword spacing. This is called "TYPRINT." The other accepted information already on punched cards or magnetic tape in fields of fixed or variable format, generating, where appropriate, symbols to enable the typesetting machine to set the selected material in a variety of type faces with appropriate uppercase characters. This is "TABPRINT." Still another option, "BCDPRINT" was developed to facilitate the input of Hollerith cards, and some relatively simple but attractive mathematical equations were developed from linearized representations, in conjunction with J. M. Gerard. One of the most valuable aspects of the book is its illustrative material, since it shows not only numerous samples of output but the coded input which was required in order to elicit that output.

It is not this reviewer's purpose to appraise Barnett's programs. It is one thing to produce output in a laboratory. It is another to arrive at timely, economic, and reliable production solutions. Barnett's work was concluded before he got to what seems to many to be the heart of the problem—namely, the development of reliable and efficient updating and correction procedures. Without these, phototypesetting by computer is clearly impractical. Work done more recently at the University of Pittsburgh (the inheritor of the Barnett Photon and programs) has placed a great deal more emphasis upon pre-typesetting revisions, although their solutions seem presently to be out of the question in terms of commercial economics.

Barnett uses the term "Computer-Aided Typesetting Process" or CATP to describe projects of all varieties in this field, and in the latter part of his book he summarizes some of the work of others. This summary, as the author admits, is far from complete. It suffices to give a flavor of the flurry of activity now being carried forward in many quarters but it is by no means a definitive compilation of such efforts. It virtually ignores the significant activity carried on in the commercial sphere by newspapers, computer hardware manufacturers, printers, typesetters, and centers such as our own. C. J. Duncan, of Newcastle University, seems to have assumed the role of official reporter of worldwide developments, and one can also derive a great deal of information from the conferences of the Research and Engineering Council of the Graphic Arts and those of American University in Washington, D. C. It appears that Professor Barnett added his brief summary merely to give perspective to his own experimental work

* *Graphic Composing Techniques*, RADC No. IDR 61–310, Final Report, Contract AF 30 (602)–2527.

and perhaps to lay the background for his concluding chapters which are concerned with "prospects," but a more comprehensive survey and appraisal would have substantially enhanced this undertaking.

Chapter 9, *"The Economics of Computer-Aided Typesetting,"* correctly enumerates most of the factors which are relevant to a consideration of costs. Unfortunately he does not weigh these factors or attempt to assign numbers to any of them. Studies have been made on the speed of keyboarding raw paper tape, for example, which permit actual keystroke costs to be projected. Realistically appraising the tremendous problem of developing appropriate typesetting software, Barnett expresses a strong preference for compiler-type programs and recommends against "tight" solutions that go to the limit of a computer's capabilities. In this respect it seems to this reviewer that he ignores the economic realities that compel commercially-oriented ventures to minimize the cost per thousand input keystrokes by striving for high operating speeds and low monthly computer rentals (with correspondingly small memory and storage areas).

The book is largely impersonal, but occasionally Barnett permits his social judgments to come to the fore. "The author has found it interesting," he states, "to try conveying the experience of participation in the development of something new to groups of senior keyboard personnel in England. He was left with the impression that these people's attitudes were conditioned by a feeling of social incompatibility with participation in such activities. Despite the fact that, in principle, educational opportunities are available in England to children of every economic background, there are very strong social factors that limit the acceptance of these in a craft or artisan environment, and when they are accepted in such an environment, a fairly strong attempt is usually made to break away from an employment that is associated with it. Interrelated factors seem to be involved in the attitude to transition from the shop floor to management, which happens only rarely in England and when it does, is regarded almost as some form of betrayal of principles rather than as a progression. England may be particularly unfortunate in this respect. The accent of speech in England carries many of the barriers to change, passed on from generation to generation, which sociologists ascribe to the color of skin in the United States."

Elsewhere, in discussing programming tactics, he touches on a peculiarly vulnerable relation between programmer and manager: "A novice who has just learned to program underestimates the complexity of a job in which he wants to use a computer, underestimates the work of writing a computer program, is ignorant of the existence of extensive programming practicalities that are not taught in programming courses and are really learned only from protracted personal experience, is highly enthusiastic and optimistic about a program he decides to write, and strongly opposes any attempts to guide or supervise his work. Quite often a situation arises in which a novice in this frame of mind is indulged by a manager who has a very casual knowledge and who is overawed by the novice's enthusiasm and jargon." He warns about the "highly personal program that may allow its author to exercise a temporary tyranny over his employer, but which may also turn into an acute embarrassment for the programmer." He comments about the "pernicious" outside influences which beset work of technological potential and warns that the "university must not be associated . . . with exaggerated claims or premature attempts at implementation." In passing, he observes that "America's great advantage is its administrative structure," and "Britain places faith in an oligarchy which is not constrained by checks and balances of the sort which exist in the United States."

The last overly-brief chapter of "conclusion" points the direction new research might take, indicates the probable trend of CATP and relates typesetting by computer to "information generating procedures," "the transformation of information," and information retrieval. "CATP'S may well augment the flood of printed pages that are produced in the cause of academic tenure, managerial competition, ego satisfaction, and the fear of making decisions that seek protection in anthropomorphizations of the digital computer. Nonsense will gain authority of the well-typeset page in much greater quantities that now. . . ."

There is much sense and no nonsense in Barnett's useful book. It is a treatise which will be of value to anyone concerned with the "knowledge explosion." Its inadequacies stem largely from the fact that we are not dealing with a science with its well-defined frame of reference, so that it is not clear to the reader what is strictly pertinent. Here technology is opening new vistas. Barnett helps us to see some of them at close hand, and with occasional glimpses of clarity, to peer somewhat farther down the road toward the future.

JOHN W. SEYBOLD
President
ROCAPPI, Inc.

10/66-3R **Mental Health Book Review Index.** Vol. 10, Whole No. 15. 1965. Compiled by the Editorial Committee and Contributing Librarians. Council on Research in Bibliography, Inc., New York. 78 pp.

This is the tenth anniversary volume of the *Mental Health Book Review Index* which now has listed a total of over 18,000 references to reviews of 3,283 monographs dealing with mental health, a field defined by the editors of the *Index* as one which "makes use of a core of knowledge contributed by the behavioral sciences and amplifies it with scientific, professional, and social ramifications and applications." The annual volume this year includes around 300 titles in alphabetical order by author and gives for every entry three or more signed reviews selected from over 200 English language journals. Titles in one year's list may be repeated later if three or more additional reviews have subsequently appeared. Though the behavioral sciences are broadly represented, the emphasis is on the psychological sciences at the center, for at least one review for every book selected must come from a journal in the latter area. The intent is to cover important topics fully, though not exhaustively.

The *Index* is produced through the combined efforts of an editorial committee chaired by Librarians Ilse Bry and Lois Afflerbach, a group of 38 contributing librarians associated with work in mental health, and a committee of specialist consultants. As would be expected from a group of this kind, bibliographic detail is full and clear.

The uninitiated might perhaps plan to use this compilation as a book selection tool, but its listings are too late to be of value in this respect except in building a retrospective collection. Its primary objective, as discussed by Dr. Bry and her co-chairmen in a series of editorial prefaces written for the annual issues. has been of a quite different nature. The aim has been, through a review of reviews, to develop a mechanism "for identifying and organizing an evolving literature in a new domain of knowledge." Monographs have been thought of as synthesizing and interpreting the research literature of this new domain and much attention has been concentrated, especially in the later years, to its multidisciplinary character.

The editorials contain some very interesting observations. Titles have been found to be under critical review for as long as five years. At the same time, less than a quarter of the books reviewed in the journals selected for indexing have received three or more reviews, although some have received 25 or more. Thus the titles which have met the criteria for listing in the *Index* have been given a collective evaluation by the scientific community. The evaluation serves not only as an assessment of research in the behavioral sciences, but as a device for showing by their absence from the review record important areas which need to be brought into focus.

There are judicious words on the subject of bibliographic style as a means of scholarly communication, on proper bibliography as a means for allowing a maximum number of correct and significant inferences from a minimum of words, and on the futility of attempting to control the literature crisis by the indiscriminate listing of all the publications in a given field rather than by separation of what is worth retrieving from what is not.

The thoughtful, scholarly approach to bibliographical problems brought out in its editorials deserves a wider audience than the special subject area and format of the *Index* customarily draws. The latest discussion, for example, is devoted to the need for an entirely new kind of classification for organizing the behavioral science literature. The evolution of the concept of the sciences of man in the 16th century into the behavioral sciences of the 20th century is traced in parallel with the development of classification schemes during the same periods. Two divergent trends in classification are noted: the anthropocentric, in which subjects are grouped in the sequence in which man might reasonably pursue them in an ordered search for knowledge, and the anthropotropic, in which man turns to himself to study the nature of man. The major library classification schemes are anthropocentric, but the behavioral sciences are essentially anthropotropic; hence the impasse in attempting to relate them satisfactorily with one another and with the collection as a whole. The editors list three major problems: (1) the need to transfer psychology from its 19th century position as part of philosophy as in the Dewey Decimal and Library of Congress classifications, (2) the integration of psychology and psychiatry now separated in all general classifications except Bliss, and (3) determination of whether psychology can be confined within a class at all. In actual fact, the first two problems would disappear if the third could be solved. The solution proposed is an open orbital system for the behavioral science literature rather than the traditional linear scheme. The psychological sciences (psychology, psychoanalysis and psychiatry) form the center of the orbit and thus integrate the anthropocentric and anthropotropic approaches. Other fields find their place in the orbit in accordance with their degree of interaction with the psychological sciences. New fields and new interdisciplinary areas can obviously be accommodated at will in such a system. The scope of the system is suggested as "knowledge about behavior in its roots and manifestations, in man and animals, in individuals, groups, and culture, and in all conditions, normal, exceptional, and pathological."

The proposal is certainly an interesting one, but certain of its assumptions require elaboration. Though the need to delimit the behavioral science literature from the total literature of man is postulated, the definition of areas of interaction is so broad that precision in separating one from the other would be difficult indeed to obtain. The orbital idea might function very well in a system designed for users primarily from the psychological area, but the practicality of combining orbital and linear schemes in one general bibliographic system which would, for example, in the area of neurochemistry give equal satisfaction to the psychologist, neurologist, and chemist is not so evident. The editors go on to state that the *Index* has for the past ten years been experimenting with the orbital organization of the behavioral science literature. This statement also needs further explanation, though the implication is that the manner in which reviews are selected would automatically insure an orbital pattern if the reviews were rearranged by subject areas instead of by author as at present.

This volume of the *Mental Health Book Review Index*, like it predecessors, is its own best witness that it has been compiled with careful thought for the time and people it serves.

Louise Darling
Biomedical Library
University of California, Los Angeles

10/66–4R **Mining, Minerals, and Geosciences.** Volume 2 of Guides to Information Sources in Science and Technology. 1965. Stuart R. Kaplan. Interscience Publishers, a division of John Wiley & Sons, New York City, 599 pp.

The aim of this book — to provide a comprehensive guide to continuing sources of information in the fields of metallic and non-metallic mining, metals, fuels, minerals, geology, geophysics, beneficiation and processing, geography, and the broad area of pure and applied earth sciences—is good; the result: somewhat less than a bull's eye, but useful nevertheless.

Part I, 445 pages, lists organizations which are sources of information about one or more of the broad subject areas covered. These sources are arranged by nine major geographical areas, starting with international organizations and then proceeding to North America, Central America and Caribbean Islands, South America, Europe, Africa, Middle East, Asia, and Oceania. Listed within each of these geographical areas are the organization name, address, telephone number, telegraph and cable address, brief description of purpose and functions, year organized, organizational structure, divisions, departments and sections, and their function, regional, branch, and district offices and addresses, membership, and publications. This information appears in alphabetical order by organizations, under each country in the geographical areas, except that Canada, United States, and United Kingdom are further subdivided into more manageable units such as government agencies; scientific organizations, institutes, and associations; and states or provinces.

Part II of the book lists in 116 pages the published literature in the fields of geography, geology, geophysics and geochemistry, glaciology, lapidary, mineralogy, mining and metallurgy, oceanography, paleontology, photogrammetry, physics, science, seismology, soil science, speleology, ceramics, coal, gas, iron and steel, and petroleum. Within these major subject areas the literature is classified first by types, i.e., abstracts, bibliographies, dictionaries, directories, handbooks, or journals, and within each of those types the publications are subdivided by geographical areas.

This directory of organizations can be useful for seekers of contacts in various parts of the world. For example, this reviewer is scheduled to pass through Calcutta, India, in connection with a visit to a dam project, and was interested to learn (pp. 393–394) that the Geological Survey of India, located in that city, is likely to have information on engineering geology for the study of dam sites — perhaps this very one under study.

Recognizing that chances for incompleteness of information are good, the book contains a tear sheet for the reader to make recommendations, omissions, and corrections which can be mailed to the publisher to the attention of the editor. In perusing the volume, I found several places where the information furnished either lacked being complete or was entirely omitted. For example, under GEOPHYSICS AND GEOCHEMISTRY, journals, USSR, the first item is:

GEOHIMIJA
Geohimija, Vorobyevskoye shosse 47—a, Moscow, USSR.

This is considerably less informative than can be found in entry 689 of a previous publication, *A Guide to the World's Abstracting and Indexing Services in Science and Technology, Report No. 102, National Federation of Science Abstracting and Indexing Services, Washington, D. C., 1963*, published under a grant from the National Science Foundation. This entry is:

GEOKHIMIYA

Akademiya Nauk USSR; order from "Akademkniga," Pushkinskaya 23, Moscow K–104 USSR Monthly; since 1956; 75 abstracts and 25 references a year to world literature; 9 rubles. geochemistry

The absence of duplicate entries or at least cross references also creates lack of clarity insofar as abstracts in geophysics and geochemistry are concerned. For example, on page 473, only one publication is listed under the category of abstracts, i.e., Geophysical Abstracts (U. S. Geological Survey) in the field, GEOPHYSICS AND GEOCHEMISTRY. However, on page 457 under GEOLOGY in the category of abstracts, *Geoscience Abstracts* is listed. The description of the latter publication states that it includes material in geophysics and geochemistry. Thus, if someone is interested in locating all the abstract publications covering geophysics, he would need to look in several places for the information.

Still on the subject of geophysics, this reviewer looked for an entry describing the abstract journal in geophysics prepared by VINITI (the institute for scientific and technical information) which he visited in Moscow as part of a UNESCO team in 1963. It was not listed in the book, even though it appears as item 1418 of the publication referred to previously. The reference is:

> *Referativnyi Zhurnal: Geofizika,* monthly; since 1957; 15,500 abstracts a year from world literature; annual author and subject indexes; 27 rubles, 60 kopecks (for organizations), 17 rubles, 25 kopecks (for individuals).

This small sampling which is limited to publications of abstracts may not be indicative, but it seems to the reviewer that a list of sources of information that purports to be comprehensive should include pertinent material available in published form elsewhere.

Three indexes are provided: (1) an Index of Geographical Areas, (2) an Index of Literature, and (3) an Index of Organizations. These add substantially to the accessibility of the information to be found in this volume, but the previously mentioned omission of "geophysics" as a subject of Geoscience Abstracts is consistent with its omission in the Index of Literature.

It is hoped that the first planned revision of this reference work to keep it up-to-date will also result in a more comprehensive volume.

Jack W. Hilf
Water Research Scientist
U. S. Dept. of the Interior

10/66–5R **Library Science Today: Ranganathan Festschrift Vol. II.** 1965. Edited by Kaula Prithvi Nath. Ranganathan Series in Library Science, 14. Asia Publishing House, New York, 832 pp. $30.00.

The second volume of this Festschrift is a Ranganathan bibliography compiled by A. K. Das Gupta. This volume contains 117 signed articles by 109 authors (some of them are only joint authors), some of whom wrote two or more papers. The authors are from the following countries, among others: Australia, Austria, Germany, Great Britain, India, Japan, Nepal, Netherlands, Norway, Pakistan, Switzerland, Thailand, and the U. S.

Among U. S. authors are James B. Childs, Robert B. Downs, Theodore A. Mueller, Jesse Shera (with James W. Perry), Ralph R. Shaw, Louis Shores, Maurice F. Tauber, and Lawrence S. Thompson.

The main groupings are: classification (general), colon classification, faceted classification, cataloguing (general), cataloguing in Japan, subject cataloguing, documentation, laws of library science, librarianship, library movement, library organization, university libraries, library administration, reference tools, social education, library education, evaluation (works), evaluation (works and life), evaluation (life), and reminiscences. Appendixes contain the list of members of the Ranganathan Commemoration Volume Committee, a list (with identification) of authors of papers, a chronology of Dr. Ranganathan's life, and the Committee's appeal for contributions. A helpful index is included.

While many of the articles deal with other subjects, the great majority of them concern Dr. Ranganathan's life and his place in librarianship. Much of the basic biographical material is repeated many times by various authors. It is generally asserted that his contributions have been very great, frequently with a touch of what one suspects reflects national pride. One wonders about statements such as the following: "He is the international figure in the Library World" (p. 532); "the Master-Architect in Library Science" (p. 550); "the Father of Library Education" and "a multi-faceted genius in Library Science" (p. 554).

On the other hand, Americans and others have joined in giving a high evaluation of Dr. Ranganathan's contributions. Shera and Perry claim that he "found librarianship little more than a bundle of techniques, a rather simple technology, and he, and his followers, have raised it to an intellectual discipline in its own right" (p. 45).

Again, Coblans asserts: "it is Ranganathan who made the first break-through in classification theory which provides possible structural techniques for handling scientific information" (p. 281).

Shaw calls him "a truly great teacher" (p. 63), and evaluates colon classification as "a milestone in the development of the intellectual process involved in the organization of the information that appears in recorded form" (p. 63).

Vickery thinks that Dr. Ranganathan's "scientific approach to classification" is "his most enduring contribution to librarianship" (p. 109).

Despite much repetition, this book is a contribution to library literature. It tells us about an outstanding figure in librarianship, and its Festschrift character has brought in several contributions on other subjects which will be missed if one regards the book as devoted solely to Dr. Ranganathan. A few of these may be mentioned: cataloguing in Japan, corporate author entry for the German Federal Republic, government and official publications in the German Democratic Republic, copying the old catalogue of the Austrian National Library, books for Norwegian seamen, bookmobile service in Hawaii, Farmington Plan for Pakistan, the future of university libraries (Downs), university library building planning, the libraries of the UN, encyclopedia making, printing and printing collections in Kentucky, early history of European periodicals, and books for children and youth in the German Democratic Republic.

The book is printed in a satisfactory format, but on a poor quality paper, and is not well-bound. A good many typographical errors got past the proofreaders. A large number of photographs add to the book's usefulness. The price of $30.00 seems quite high.

Dr. Luther H. Evans
Director, International and Legal Collections
Columbia University

10/66–6R **Use of Mechanized Methods in Documentation Work.** 1966. Herbert Coblans, ASLIB, London. 89 pp.

This is an excellent tutorial and critical review of the state of mechanization in libraries, documentation centers, and other information handling activities. From its introduction on, it emphasizes the difference between what it calls the "whole hog" approach to mechanization and the "housekeeping" approach — the one trying to put everything into the computer, the other adopting a less mechanized mix. The report consists of an introduction, then three parts covering different areas of mechanization, and finally, three appendices.

Part I emphasizes the more or less clerical areas of mechanization — the housekeeping functions such as catalog production, serial records maintenance, circulation, and acquisition accounting and control. Part I concludes with a discussion, pro and con, of the role of mechanization in these areas. This discussion can be summarized as follows: The principal advantage of such mechanization in the likelihood of improved efficiency and effectiveness, but it is hard to confirm just how much there will be because staff is transferred to other service activities, present costs are not well established and are not really comparable anyway, the rigid limitations imposed by mechanized operations change some of the qualitative characteristics of operation. In general, however, the evaluation is a positive one.

Part II discusses reference and document retrieval (with the prototypical example being Medlars). It reviews several of the approaches to it — hardware, software, and total systems — including permuted indexes, image storage, TIP, SDI, etc. It also concludes with an evaluative discussion, based primarily on Medlars. Since both NLM and its experimental sub-centers at Colorado and UCLA have presented a continuing qualitative and quantitative evaluation of this system, it constitutes an excellent basis for operational evaluation. (The recent article* in the Bulletin of the Medical Library Association is a model for anyone reporting on such systems.)

Part III discusses the application of computers to data retrieval, but so briefly that its value lies solely in having recognized that data banks exist.

Appendix 1 is a relatively up-to-date bibliography; Appendix 2 describes existing "non-conventional" systems in the United Kingdom, and Appendix 3 is a brief listing of some equipment available in the United Kingdom. There is an index, primarily to names.

Dr. Robert M. Hayes
Professor
School of Library Service
University of California at
Los Angeles

* Rogers, Frank B., "MEDLARS operating experience at the Universtiy of Colorado," Bulletin of the Medical Library Association, 54(1):1-10, Jan. 1966.

ADI Chapters and Chapter Secretaries

CENTRAL OHIO CHAPTER
Mrs. Arleen N. Somerville
Chemical Abstracts Service
Ohio State University
Columbus, Ohio 43210
614-293-6586

CHICAGO CHAPTER
Mrs. Barbara N. Yanick
Librarian, NALCO Chemical Co.
6216 W. 66th Place
Chicago, Ill. 60638
312-PO7-7240, ext. 472

DELAWARE VALLEY CHAPTER
Miss Nellie A. Medzadour
Wyeth Laboratories
Box 8299
Philadelphia, Pa. 19101
215-MU8-4400, ext. 678

INDIANA CHAPTER
Mr. Asa N. Stevens
6157 E. St. Joseph Street
Indianapolis, Indiana 46219
317-357-6460

LOS ANGELES CHAPTER
Myra Grenier
Aerojet General Corp.
P.O.Box 296
Azusa, Calif.
213-334-6211, ext. 5166

METROPOLITAN NEW YORK CHAPTER
Miss Nanette Farley
T. J. Watson Research Center
IBM Corp.
P.O. Box 218
Yorktown Heights, New York
914-WG5-2037

NEW ENGLAND CHAPTER
Miss Virginia Valeri
Arthur D. Little, Inc.
15 Acorn Park
Cambridge, Mass.
617-864-5770

NORTHERN OHIO (CLEVELAND) CHAPTER
Miss Helen Skowronska
Sherwin-Williams Co.
P.O. Box 6027
Cleveland, Ohio 44101
216-TO1-7000

PITTSBURGH CHAPTER
Mr. James Brandt
ALCOA Research Labs.
Box 772
New Kensington, Pa.
412-337-6541

POTOMAC VALLEY CHAPTER
Mrs. Mary Herner
Herner & Co.
2431 K Street, N.W.
Washington, D.C. 20037
202-965-3100

SAN FRANCISCO CHAPTER
Miss Marilyn Johnson
Shell Development Corp.
1400 53rd Street
Emeryville, Calif. 94608
415-OL3-2100

SOUTHERN OHIO CHAPTER
Miss Marie L. Koeker
2445 Fairport Avenue (home)
Dayton, Ohio 45406
513-235-3419 (office)

SOUTH TEXAS CHAPTER
Mr. Doug Yauger
7107 Augustine
Houston, Texas
713-PR4-1269

UPSTATE NEW YORK CHAPTER
Mr. Herbert Ohlman
Xerox Corp.
Box 1540
Rochester, N.Y. 14603
716-TR2-2000, ext. 22158

Index to American Documentation
Volume 16 and Volume 17 (Nos. 1, 2, 3 only)

This is a deep, informative, subjective index to ideas and concepts contained in papers, brief communications and letters to the editor, produced with the IBM Selectric typewriter.

Authors and subject headings are in one alphabetical listing. Page numbers refer to pages in Vol. 16, unless underlined, in which case they refer to Vol. 17. *Items from letters to editors* and authors and titles of reviewed books are in different type faces. User comments are invited.

Isaac D. Welt
Associate Editor.

AMERICAN DOCUMENTATION

VOLUME 17

1966

Published by

THE AMERICAN DOCUMENTATION INSTITUTE

CONTENTS

No. 1, January 1966

Brief Communication

No. 2, April 1966

Brief Communication

No. 3, July 1966

No. 4, October 1966

Documentation Abstracts

. . . . . is a joint publication under the auspices of the American Documentation Institute and the Chemical Literature Division of the American Chemical Society.

. . . . . represents combined coverage of the former Literature Notes section of American Documentation, the ACS Division of Chemical Literature Annotated Bibliography, and the former coverage of Documentation Digest.

. . . . . will issue quarterly — February, May, August, and November of 1966. Each issue will contain corporate and author indexes; subject indexes will be available on a schedule to be determined.

Subscriptions are sold on a calendar year basis — $8.00 per year.* Return the coupon below. Payment with your order is requested.

* Members of the American Documentation Institute will receive the first year's subscription free.

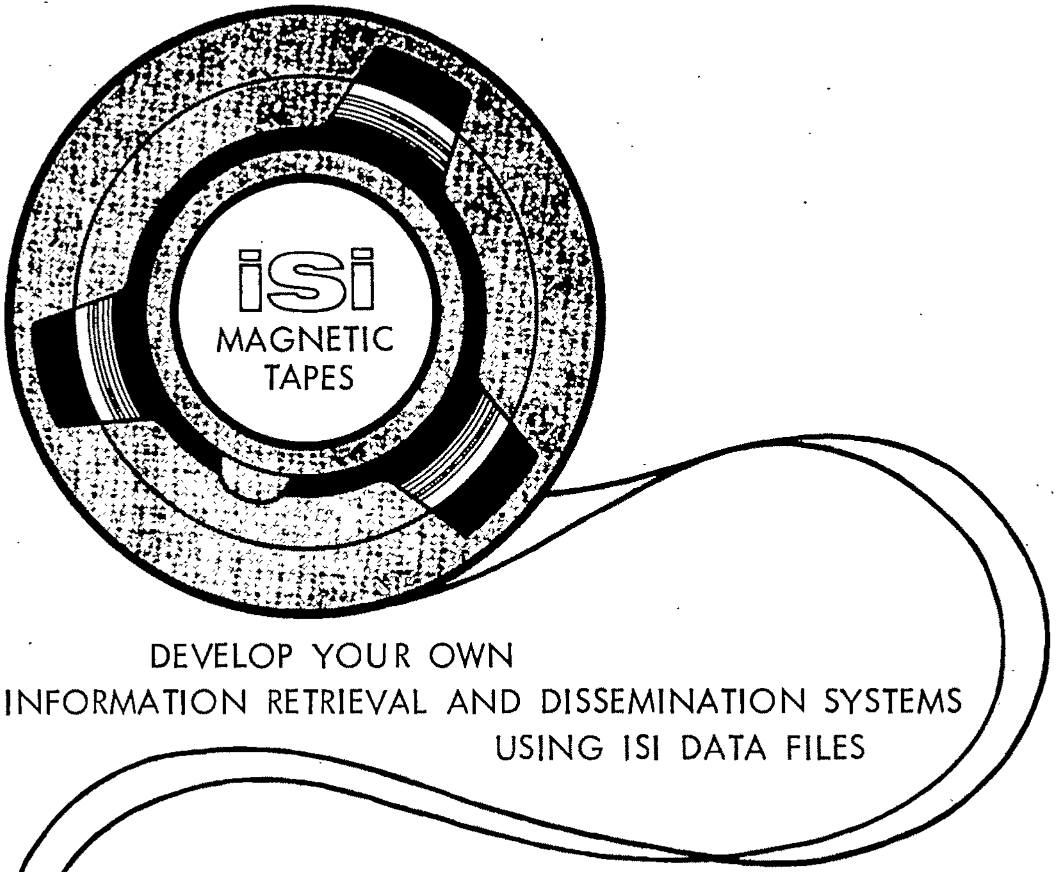
DOCUMENTATION ABSTRACTS—Please enter my subscription for one year commencing with the February 1966 issue. At $8.00 per year, payment is enclosed □ bill me □

Name____________________ Title_______________

□ Business
Address □ Home ________________________

City ________________ State____________ Zip______

Your Firm Name ____________________________



DEVELOP YOUR OWN
INFORMATION RETRIEVAL AND DISSEMINATION SYSTEMS
USING ISI DATA FILES

ISI is now making available to scientific and technical information centers, interdisciplinary magnetic tape files of the current literature of science and technology. These are the most comprehensive files available anywhere in the world -- and you pay only for the information you use.

These data files can be used for SDI systems, alerting publications, KWIC indexes, retrospective searching, etc.

If you would like to explore the possibilities of these unique weekly tapes, please contact us for information. Write Dept. 07-5.

another service of isi

INSTITUTE FOR SCIENTIFIC INFORMATION 325 Chestnut St Philadelphia Pa 19106 USA